REMARKS

The Office Action of May 14, 2004 presents the Examination of claims 1-10, 16-23 and 28-30. These claims are not amended and remain pending.

Substitute Specification Provided

The Examiner objects to the specification as presenting an overly detailed Summary of the Invention portion that amounts to merely a restatement of the language of the claims. The Examiner's objection is addressed by provision of a substitute specification deleting that portion of the specification.

The Examiner also objects to informalities of a specification "filed April 14, 2000," related to incongruity between the description of sequences presented in the Sequence Listing and sequence listing identifier numbering in the instant specification. It appears that this objection arises due to confusion by the Examiner of the specification in the instant application and the specification of record in the co-pending application 08/992,914. Applicants note first that there was no specification filed April 14, 2000 in the present application. Applicants further note that the present specification does not include any Example 10 or 11 and furthermore that the Example 6 appears on page 45, not on page 37. Thus, it appears that the Examiner's reasons for objecting to the specification are not correct.

Applicants have incorporated the amendments of the Sequence Listing identifiers from the Preliminary Amendment of April 29, 1999 into the attached substitute specification.

No new matter is added by the substitute specification. A "track-changes" version of the substitute specification is attached hereto for the Examiner's review.

Rejection under 35 U.S.C. § 112, second paragraph

Claims 1-10, 16-23 and 28-30 are rejected under 35 U.S.C. § 112, second paragraph, for allegedly being indefinite in the recitation of incorrect Sequence Listing identifier numbers. Applicants believe that the Examiner bases this rejection upon reading of the wrong specification and therefore the rejection should be withdrawn.

Rejection under 35 U.S.C. § 101 and 35 U.S.C. § 112, first paragraph

Claims 1-10, 16-23 and 28-30 stand rejected under 35 U.S.C. § 101 and 35 U.S.C. § 112, first paragraph, for alleged lack of description of a utility or assertion of a known utility for the claimed invention. Claims 1-10, 16-23 and 28-30 are also rejected under 35 U.S.C. § 112, first paragraph for alleged lack of adequate written description of the invention. These rejections are respectfully traversed. Reconsideration and withdrawal thereof are requested.

The Examiner takes a position that there is no evidence of record that the claimed nucleic acids encode a protein having raffinose synthase activity. The Examiner asserts that mere homology to a known raffinose synthase gene is not sufficient, as there are other genes known that have higher degrees of homology to some raffinose synthase gene, yet encode proteins having other activities. In particular, the Examiner has compared the *S. affinis* stachyose synthase to the *S. cuminis* raffinose synthase and found a degree of homology of 50% and with the *P. sativum* raffinose synthase and found a degree of homology of 51%. Applicants note that in their prior response, a sequence comparison was provided as Table 2; these same comparisons are in Table 2 and a degree of homology of only 43% is found for both.

The discrepancy in the degree of "homology" in the data set provided by the Applicant and by the Examiner is due to differences in the computer program used to analyze the data. Table 2 of Applicants' previous response shows overall sequence homologies(%) between raffinose synthases (RFSs), imbibition protein (SIP) and stachyose synthases (STSs). Applicants' homology data were calculated using Global Alignment (the alignment of sequences over their entire length) produced by the CLUSTAL sequence analysis program. The CLUSTAL program uses the algorithm of Wilbur and Lipman (see the attached Exhibit 1, a description of sequence

analysis programs found at http://www.rfcgr.mrc.ac.uk/embnet.news/vol2_1/align.html).

On the other hand, the Examiner's analysis was conducted using the BLAST program, which is a Basic Local Alignment Search Tool. The BLAST program uses the BLAST algorithm and makes local alignments (the alignment of some portion of two sequences) in order to search similarities of sequences. BLAST is a local alignment program, and does not make global alignments between sequences to calculate total percent homologies. (See, page 3, the 9th paragraph of attached Exhibit 2, http://www.ncbi.nlm.nih.gov/BLAST/blast_FAQs.shtml). Moreover, "identities" values output in a BLAST search report are different from "homology" values. That is, "homology" means "similarity attributed to descent from a common ancestor", while "identity" means "the extent to which two sequences are invariant" (See, pages 2-3 of attached Exhibit 3, http://www.ncbi.nlm.nih.gov/Education/BLASTinfo/glossary2.html). Thus, different values and scores calculated using different algorithms are based on different standards or criteria, and therefore it is not reasonable to discuss homology and similarity of sequences only by directly comparing such different values and scores.

The attached Table 3 shows the identities obtained using a similarity search using the BLAST program for the amino acid sequences of RFSs, SIP and STSs shown in Table 1 (attached to

Applicants' previous response). Among Sc-02, Sc-03, Sc-04 and Sc-05, the identities were obtained by searching a patent database with default parameters, using the amino acid sequence of each protein as the "query", and using "Protein query vs. translated database (tblastn)" of the NCBI BLAST program. Also, other identities were obtained by searching the non-redundant database with default parameters, using the amino acid sequence of each protein as the "query", and using "Protein-protein BLAST (blastp)" of the NCBI BLAST program.

The identities between RFSs and SIPs are about 40%. The identities between RFSs and STSs range from about 40% to about 50%. On the other hand, the identities among RFSs are 60% or more. The identities among STSs are also 60% or more. That is, the identities among RFSs or the identities among STSs are higher than the identities between RFSs and SIPs or the identities between RFSs and STSs. Thus, based on the results of analyses by BLAST program, RFSs, SIPs or STSs can be distinguished. Applicants note that the conclusion reached from this analysis is consistent with the conclusion reached using the CLUSTAL analysis provided in their previous paper.

Applicants have also previously provided the Declaration of Dr. Watanabe to address the issues of adequacy of the written description and disclosure of utility. The Examiner has dismissed the Watanabe Declaration, stating that it is unclear what nucleic acid was used in the examples in the Declaration. The Watanabe

Declaration describes on its page 2, at the beginning of the item "Experiments":

> The vector BjRS-Sac(+)-121 having the mustard raffinose synthase gene of the present invention in the expressible direction (i.e. sense direction) and the vector BjRS-Sac(-)-121 having the mustard raffinose synthase gene of the present invention in the reverse direction (i.e. antisense direction), which are the same as obtained in Example 8 of the present specification, were used for the transformation of tobacco (Nicotiana tubacum).

The Examiner should take note that Example 8 of the originally filed specification (page 47, lines 13-16) describes "the nucleotide sequence of SEQ ID NO:6 and SEQ ID NO:8 were determined from mustard (Brassica Juncea) and rapeseed Wester (Brassica napus), respectively". Therefore, the mustard raffinose synthase gene used in Watanabe Declaration is the gene from the mustard Brassica Juncea.

The Examiner also dismisses the Watanabe Declaration as presenting only a single example of a functional raffinose synthase gene. The Examiner's argument that Dr. Watanabe establishes functionality of only one gene is not persuasive. Dr. Watanabe's Declaration provides evidence of functionality in support of an assertion of utility that cannot be dismissed merely because of the number of examples it sets forth. Dr. Watanabe's Declaration provides evidence that the asserted utility for the invention is indeed correct; any "rebuttal" by the Examiner must be supported by

evidence, not a mere assertion that not enough examples have been provided.

The Examiner "rebuts" Applicants' arguments about the sufficiency of the disclosure in the specification by suggesting that the sequences mentioned in Applicants' arguments are not those set forth in the sequence listing. Applicants note the Examiner's confusion of the instant application with a co-pending application and urge that he revisit Applicants' previous argument.

The Examiner should note that it is Applicants' burden to establish adequacy of written description and utility only by the preponderance of the evidence. Applicants submit that the evidence of record firmly establishes 1) that raffinose synthase genes as a genus are more closely related to each other than to stachyose synthase genes or imbibation protein genes suggested by the Examiner as establishing lack of utility, and therefore the skilled artisan can distinguish a raffinose synthase gene from either of a stachyose synthase gene or an imbibation protein gene by proper homology analysis and 2) that a gene identified as a raffinose synthase gene by homology analysis is most likely to actually have the biochemical activity of a raffinose synthase gene. (This latter being established by the Watanabe Declaration at least.)

Accordingly, the utility of the instantly claimed invention is established and the rejection of claims 1-10, 16-23 and 28-30 stand rejected under 35 U.S.C. §§ 101 and 112, first paragraph,

for alleged lack of proof of an asserted utility should be withdrawn.

The adequacy of the written description of the invention is established by the comparison of the specification with the Guidelines for examination of written description in Applicants' previous response. The Examiner is requested to reconsider his position and arguments upon reading of Applicants' arguments together with the specification of the present application, not the specification of the co-pending application.

Obviousness-type double patenting

Claims 1-3, 16-23 and 28-30 are rejected under the judicially created doctrine of obviousness-type double patenting over claims of the co-pending application "08/992,766". First, the Examiner is requested to confirm that the co-pending application is in fact 08/992,914. Second, the Examiner is requested to hold this rejection in abeyance until either this application or the '914 application is allowed, at which time an appropriate response in the form of either arguments distinguishing the invention or a terminal disclaimer will be filed in the application that remains under examination.

Conclusion

Applicants respectfully submit that the above remarks and/or amendments fully address and overcome the rejections of record.

The present application is in condition for allowance. The Examiner is respectfully requested to issue a Notice of Allowance indicating that claims 1-10, 16-23, and 28-30 are allowed.

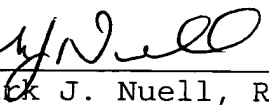
Should there be any outstanding matters that need to be resolved in the present application, the Examiner is respectfully requested to contact Mark J. Nuell, Ph.D. (Reg. No. 36,623) at the below-listed number.

Pursuant to the provisions of 37 C.F.R. §§ 1.17 and 1.136(a), the Applicants hereby petition for an extension of three (3) months to November 14, 2004, in which to file a reply to the Office Action. The required fee of $980.00 is enclosed herewith.

If necessary, the Commissioner is hereby authorized in this, concurrent, and future replies, to charge payment or credit any overpayment to Deposit Account No. 02-2448 for any additional fees are required under 37 C.F.R. § 1.16 or under 37 C.F.R. § 1.17; particularly, extension of time fees.

Respectfully submitted,

BIRCH, STEWART, KOLASCH & BIRCH, LLP

By 
Mark J. Nuell, Reg. No. 36,623

P.O. Box 747
Falls Church, VA 22040-0747
(703) 205-8000

GMM/DRN:kdm
0020-4559P

Attachments:
1. Table 3: Homologies (%) among RFS, SIP, and STS enzymes;
2. Exhibits 1-3;
3. Substitute specification (track changes copy) and
4. Substitute specification (clean copy)

Table 3

Code	Sc-02	Sc-03	Sc-04	Sc-05	PsRFS	Aj-05	HvSIP	AmSTS	PsSTS-1	PsSTS-2	SaSTS	VaSTS
amino acids	800	783	781	777	798	784	757	868	853	847	863	857
Sc-02		61	75	61	88	63	41	52	50	50	49	50
Sc-03			65	63	62	66	40	49	39	39	49	49
Sc-04				67	75	65	38	39	50	43	51	52
Sc-05					60	70	40	51	51	51	52	50
PsRFS						65	40	51	49	49	51	50
Aj-05							39	49	50	50	50	50
HvSIP								38	38	38	38	38
AmSTS									63	64	81	62
PsSTS-1										96	63	72
PsSTS-2											64	72
SaSTS												64
VaSTS												

ALGORITHMS FOR MULTIPLE SEQUENCE ALIGNMENTS

Guy Bottu,
BEN, The Belgian EMBnet node.

Introduction.

In a previous issue of embnet.news, we considered the alignment of pairs of sequences and the search for similar sequences in databanks. We now turn our attention to multiple sequence alignments.

If you have several similar nucleic acid or protein sequences it is often useful to align corresponding bases or amino acids in columns. For instance, you might wish to group bases or amino acids that occupy similar positions in the three-dimensional structure which exercise similar functions or that have evolved by substitution from the same base or amino acid in an ancestral sequence. In the latter case you might also like to construct a phylogenetic tree.

1. Global alignments.

The Needleman and Wunsch algorithm for finding the best global alignment of two sequences can readily be extended to multiple sequences. The problem is that the time the computer needs for such a job is roughly proportional to the product of the sequence lengths. So, if aligning two sequences of 300 positions takes 1 second, aligning 3 sequences takes 300 seconds and aligning 10 sequences would take 300^8 seconds, which is longer than the lifetime of the universe!

Since searching for a best global alignment using a rigorous algorithm is not realistic for more than three sequences, a number of strategies have been developed to carry out a multiple global alignment in a reasonable amount of time with a reasonable chance of finding the best alignment. The GCG program pileup first aligns all possible pairs of sequences according to Needleman and Wunsch (for n sequences, this makes n*(n-1)/2 alignments). Then it uses the pairwise similarity scores to construct a tree using the UPGMA method (see below). Finally, this tree serves as a guide for a progressive multiple alignment starting from the tips. Once two sequences have been aligned, their relative alignment is no longer changed. Clusters of previously aligned sequences are treated as a linearly weighted profile when they are subsequently aligned with another sequence or another cluster.

Other approaches include:

- The very popular CLUSTAL program differs only from pileup in that it performs the initial pairwise alignments using the fast algorithm of Wilbur and Lipman. CABIOS 8:189 (1992). References
  you can obtain versions of CLUSTAL for UNIX and for VAX
- Starting with a search for words of n bases or amino acids that are common between the sequences. An example is Martin Vingron's program MALI. CABIOS 5:115 (1989). References.
  MALI is not distributed freely but may be obtained from its author Martin Vingron (vingron@embl-heidelberg.de)
- PIMA uses pattern-matching, rather then profile matching, while making the progressive alignment. PNAS 87:118 (1990) References
  PIMA can be obtained for UNIX and for VAX
- Building a phylogenetic tree, using a more elaborate algorithm, as the sequences are progressively aligned. An example is Jotun Hein's program TreeAlign. Meth.Enzymol. 18:626(1990)
  TreeAlign can be obtained for UNIX and VMS from the same address as given for Clustal (see above)
- Making the best multiple alignment in a limited area of alignment space. This can only realistically be performed with eight to ten sequences.

2. Local alignments.

There are cases where sequences share a similar region but are otherwise completely different. Take, for example, the amino acids in the active site of an enzyme or transcription factor binding sites in a DNA sequence. To handle these cases local multiple alignment algorithms have been developed. Usually they only look for ungapped alignments thereby avoiding the problem of choosing the optimal gap penalty. Two such programs have been developed at the NCBI :
MACAW by Schuler, Altschul and Lipman first tries to find high scoring segment pairs (HSPs) for each possible pair of sequences using the BLAST algorithm (with the sensitivity set high). It then assembles overlapping HSPs into blocks. An interesting feature of MACAW is that it does not try to align all sequences, but can pick out only those that share similar regions. Proteins 9:180 (1991). References
There are versions of MACAW obtainable for the PC under Windows and for the Mac.
The MACAW distribution also contains Gibbs (see below) and a pattern searcher.

The Gibbs sampler algorithm involves iteratively making a profile with stretches of n bases or amino acids, selected from the sequences, and then searches this profile against one of the sequences. The result of the search is used to weight the selection of the stretches at the next run. A drawback is that the user must choose the width n and the number of elements in each sequence and thus must have a certain idea of the outcome, or run the program several times. An interesting feature is that the Gibbs sampler algorithm avoids the choice of an externally added scoring scheme since it derives the highest scoring profile, in a

self-consistent manner, from the data. Science 262:208 (1992). References .
Gibbs for is available for UNIX.

3. blast3.

It is also worth mentioning the program blast3. This searches a protein against a protein databank using the BLAST algorithm (with the sensitivity set high) and then makes threefold alignments between the query sequence and each possible pair of databank sequences that have been found. Only the statistically significant threefold alignments which are made from three nonsignificant pairwise alignments are retained. blast3 is useful in finding proteins that share a region of only weak similarity. Occasionally it can show that a query sequence makes the bridge between two databank sequences whose relationship had not yet been suspected.
You can look at the Manual.
It is possible to access a BLAST (including blast3) server at the NCBI, either through WWW or with a specific blast Internet client that you can install on your computer. More INFO is available.

4. phylogenetic trees.

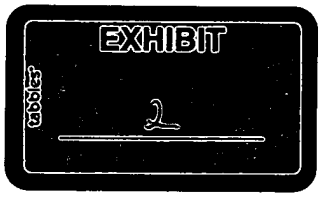
Ideally a researcher would like to have a black box in which to throw sequences and get out a fully annotated phylogenetic tree. This is, however, not possible for two reasons. First, an algorithm that considers all possible multiple sequence alignments and then, for each alignment, all possible phylogenetic trees and picks out the best one, would take too much time. That is why most phylogenetic programs work on previously aligned sequences. Second, the result is always strongly influenced by the criteria that are used to define the best tree. Phylogenetic analysis will be the subject of a separate column in a later issue of embnet.news. However, a few remarks seem appropriate here. There are three main kinds of tree building methods: distance matrix, maximum likelihood and parsimony.

Distance matrix methods first estimate the pairwise distances between the sequences (which means that the information in the alignment of two sequences is reduced to one number) while the other methods construct many trees from all the information in the multiple alignment and decide which is best.

The simplest distance based method is UPGMA (unweighted pair-group method using arithmetic averages) which involves iteratively taking together the two sequences that have the shortest distance from each other, placing them at the end of branches on a node of the tree, and replacing their distances from the other sequences by an average value.

The guide tree used by pileup and CLUSTAL should never be used to infer phylogeny! It has been derived from the distances between pairwise aligned sequences and these distances are not necessarily the same as the distances between sequence pairs taken from the multiple sequence alignment.

Go to: previous article – next article – Table to contents



Reference 2

NCBI BLAST

PubMed Entrez BLAST OMIM Taxonomy Structure

BLAST Frequently Asked Questions (FAQ)

Info
- FAQs
- News
- References
- NCBI Contributors

Education
- Program selection guide
- Tutorial
- URL API guide

Download
- Databases
- Documentation
- Executables
- Source code

Support
- Helpdesk
- Mailing list

Tips and Hints:

Troubleshooting:

Tips and hints

Q: Which BLAST program should I use?

You have many choices to make between different BLAST programs and databases. Some of these choices are better for answering some questions then others. We have created a selection chart to help you make the decision of BLAST program for the question you are asking. This is the "BLAST Program Selection Guide".

Q: How can I search a batch of sequences with BLAST?

There are three options for "Batch" BLAST searches:
1) Web MegaBLAST EST analysis tool: This program is optimized for aligning nucleotide sequences that differ slightly as a result of sequencing or other similar "errors". MegaBLAST is good for scanning a large number of EST type sequences (about 500 kb in length) against large database in search of the closest matches. You can import a file EST sequences in FASTA format or as a list of GenBank accessions or/GIs and have them compared to the BLAST databases. The default is an easily reviewable Hit Table format, although you can download and save the results in Standard pairwise HTML or any of the other result output options. MegaBLAST is available from the BLAST web page, the standalone BLAST executables, or via the network BLAST client (see below).

2) Standalone BLAST executables:The Standalone BLAST executables are command line programs which run BLAST searches against local downloaded copies of the NCBI BLAST databases. The programs will handle either a single large file with multiple FASTA query sequences, or you can create a script to send multiple files one at a time. The executables are available for a wide variety of platforms, including many "flavors" of UNIX (LINUS, Solaris, etc.) Windows PC and even Mac OSX.

The Standalone executables are available at the anonymous FTP location: ftp://ftp.ncbi.nih.gov/blast/executables/ There is information on the Standalone BLAST executables available in the README file at ftp://ftp.ncbi.nih.gov/blast/documents/blast.txt which is also bundled with the downloaded binaries.

3) BLAST Network Client 'blastcl3':The BLAST 2.0 Network client will allow you to submit a single file of FASTA sequences over an internet connection to the NCBI BLAST databases. You submit searches through the client to the NCBI servers and do not need to download the database locally. The BLAST Network client executables are located at: ftp://ftp.ncbi.nlm.nih.gov/blast/executables/ There are blastcl3 executables for various UNIX platforms, PC Windows and Macintosh.

Q: How can I write a program to submit jobs to NCBI's BLAST servers?

By using the URLAPI. Documentation also available in postscript and PDF.

Q: How can I limit my BLAST search based on Organism?

The option to limit a search to organism and even taxonomic classification is part of the "Limit by Entrez Search" option on most standard BLAST search pages. There is a pull down menu to select the most common organisms found in GenBank and also a field to input the species name, or classification (example: "eubacteria"). Using this option will cause your query sequences to be compared only to sequences in our databases from that organism.

There are also several "specialized" BLAST Pages devoted to different organisms on the main BLAST web page.

How can I limit my search to a subset of database sequences?

You can use the "Limit by Entrez Search" option found on most Standard BLASTR search pages to run an Entrez search and have your query sequence compared to the resutls of this search. For example, if you wanted to limit you search to all phosphorylase sequences from mouse you could enter the following valid Entrez search strategy in the Limit by Entrez field of the BLAST search page: phosphorylase AND "Mus musculus"[Organism]

Q: Is it possible to search for a motif or pattern with BLAST?

There are two general approaches to this type of questions. First do you wish to find if motifs exist in your query sequence, or do you have a known motif and wish to find other protiens or nucleotides with this motif?

In the first case, finding motifs in your query sequence can be done for proteins using the CDD (Conserved Domain Database) and CDART (Conserved Domain Architecture Retrieval Tool) tools. CDD allows you to compare your protein to an database of alignments and

profiles representing protein domains conserved in molecular evolution as well as 3-dimensional protein structures in the MMDB database. These tools use popular protien motif databases, PFam (http://pfam.wustl.edu/) and Smart (http://smart.embl-heidelberg.de) in addition to the MMDB database.

For conditions of the second case if you have a known motif and wish to identify other proteins with this motif you can use PHI-BLAST. PHI-BLAST searches take a motif pattern and protein sequence as input and then compares these to the NCBI protein databases looking for other proteins which contain conserved regions similar to the motif entered.

For nucleotides it is only possible to search with short query sequences representing your motif or region of interest with the Nucleotide BLAST "Search for short nearly exact matches" service from the main BLAST web page. This can find other sequences whicvh contain similar nucleotide patterns. however there are no database of nucleotide patterns which can identify patterns in your nucleotide query sequence.

You may also be interested in checking out other molecular biology web sites, such as those mentioned in the Other Molecular Biology Resources section at the end of this FAQ, for motif searching software.

Q: How do I perform a similarity search with a short peptide/nucleotide sequence?

There is a special page with pre-set parameters for searching with short sequences. You can access this page by clicking the "Search for short nearly exact matches" link on the main BLAST web page.

Essentially for these searches, the Expect value has been increased and the word size decreased to optimise for short hits which generally score a large E value require smaller word sizes to initiate formation of the HSP for extension. In addition, for proteins, the matix "PAM30" becomes the default which optimises hits to smaller sequences which have a lower percentage of evolutionary drift in general.

Q: Can I use BLAST to compare to two or more sequences in a multiple sequence alignment?

You can use the BLAST 2 Sequences service to compare two nucleotide or two protein sequences against each other using the Gapped BLAST algorithm. The this will allow you to perform a BLAST search between the two sequences allowing for the introduction of gaps (deletions and insertions) in the resulting alignment. Remember that BLAST is a "local" alignment program and does not make global alignments between sequences to calculate total percent homologies.

To compare one sequence against a specific sequence or set of sequences, you can also use a separate multiple sequence alignment program. There are many such software tools available to do this. You may also be interested in checking out other molecular biology web sites, such as those mentioned in the Other Molecular Biology Resources section at the end of this FAQ.

Q: What is the Expect (E) value?

The Expect value (E) is a parameter that describes the number of hits one can "expect" to

see just by chance when searching a database of a particular size. It decreases exponentially with the Score (S) that is assigned to a match between two sequences. Essentially, the E value describes the random background noise that exists for matches between sequences. For example, an E value of 1 assigned to a hit can be interpreted as meaning that in a database of the current size one might expect to see 1 match with a similar score simply by chance. This means that the lower the E-value, or the closer it is to "0" the more "significant" the match is. However, keep in mind that searches with short sequences, can be virtually indentical and have relatively high EValue. This is because the calculation of the E-value also takes into account the length of the Query sequence. This is because shorter sequences have a high probability of occurring in the database purely by chance. For more details please see the calculations in the BLAST Course.

The Expect value can also be used as a convenient way to create a significance threshold for reporting results. You can change the Expect value threshold on most main BLAST search pages. When the Expect value is increased from the default value of 10, a larger list with more low-scoring hits can be reported.

Q: What is low-complexity sequence?

Regions with low-complexity sequence have an unusual composition andthis can create problems in sequence similarity searching (Wootton & Federhen, 1996). Low-complexity sequence can often be recognized by visual inspection. For example, the protein sequence PPCDPPPPPKDKKKKDDGPP has low complexity and so does the nucleotide sequence AAATAAAAAAAATAAAAAAT. Filters are used to remove low-complexity sequence because it can cause artifactual hits (please also see Q: After running a search why do I see a string of "X"s (or "N"s) in my query sequence that I did not put there?)

In BLAST searches performed without a filter, often certain hits will be reported with high scores only because of the presence of a low-complexity region. Most often, this type of match cannot be thought of as the result of homology shared by the sequences. Rather, it is as if the low-complexity region is "sticky" and is pulling out many sequences that are not truly related.

Other Molecular Biology Resources:

The on-line BLAST Course was written by Dr. Stephen Altschul and discusses the basics of the Gapped BLAST algorithm. In addition the full text of the 1997 Nucleic Acids Research paper "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs" is also available on-line.

Other links:

European Bioinformatics Institute (EBI) BioCatalog
Indiana University IUBio Archive
Sequence manipulation site

Troubleshooting

Q: Causes for "No significant similarity found".

Below are several reasons that a BLAST search can result in the "No significant similarity found" message.

Short Sequences: There is a special BLAST optimized for searchig with small sequences.

Go to the main BLAST web page and select the "Search for short nearly exact matches" link for Nucleotide – Nucleotide or Protein Protein sections.

Filtering: BLAST filters regions of low-complexity (for a description of low-complexity see "What is low-complexity sequence?" below). If your sequence contains large regions of "low complexity" it may not significant hits to the database. You can turn off filtering by setting the "Filter" option to "None" using the pull down tab.

Query Format: Another reason you may see the "No Significant Similarity found" message is using the wrong type of sequence in your search.

1) Accession/GI Number or FASTA. Check that you have the Input Data set to the correct format for your Query. Set the pull down menu to "Accession number or Gi" to search with GenBank accession numbers or Gi numbers. Set to FASTA for raw amino acid or nucleotide sequences. For more information on FASTA format, click here.

2) Sequence type and Program combination. You can search with an amino acid query sequence using the blastp and tblastn programs. With nucleotide query sequences you can use blastn, blastx, and tblastx. Please note that tblastx program cannot be used with the nr database on the BLAST Web page.

For more information on the BLAST programs, click here.

Q: Why does my search timeout on the BLAST servers?

Certain combinations of BLAST searches with large sequences against large databases can cause the BLAST servers to timeout. This has to do with a limit on the server CPU's which prevents sequences which generate many HSPs from hoarding server resources.

However there are some things you can do to prevent timeout and generate results from large sequences.

– Some sequences contain large regions of ALU repeats. In this case you can select the "Human Repeat" filtering option on the main BLAST search page. This will mask repeat regions which generate a large number of biologically uninteresting hits to the databases.

– Increase the Word Size to 20 – 25. With a default Word Size of 7, the BLAST algorithm finds initial HSPs of 7 bases in length and begins extension of these from either end. In a large sequence this can generate 100's of initial HSPs between the query sequence and even a single large genomic sequence in the databases. Increasing the Word Size to 25 makes the initial HSP smaller, limiting the number small initial fragments to be extended.

– Decrease the Expect value to 1.0 or lower. Many hits from large sequences are to many small fragments in the database. The expect value for these searches is such that decreasing the expect value will eliminate these results, and concentrate on results which are more likely to contain large coding regions and genomic fragments.

If you are still seeing a "timeout" error message after making the above changes, please contact blast-help@ncbi.nlm.nih.gov with the RID of your search.

Q: Why do I get the message "ERROR:BLASTSetUpSearch: Unable to calculate Karlin-Altschul params, check querysequence" ?

This will happen if your entire query sequence has been masked by low complexity filtering. You will need to turn filtering off to get hits. For further information on filtering, please read the sections of the BLAST FAQs on Q: What is low-complexity sequence? and also Q: After running a search why do I see a string of "X"s (or "N"s) in my query sequence that I did not put there?

Q: Why do I get the message "ERROR: Blast: No valid letters to be indexed"?

You may have accidentally entered an accession number in the search box without changing the input selection from "Sequence in FASTA format" to "Accession or gi". You will also see this error message if too many ambiguity codes (R,Y,K,W,N, etc. fornucleotides) are present in your query sequence. Although BLAST allows ambiguity codes, be aware that these will always contribute a negative score in nucleic acid searches. Thus, sequences such as degenerate PCR primers with ambiguity codes maynot find any significant hits even though they may be designed from sequences that are present in the database.

Q: After running a search why do I see a string of "X"s (or "N"s) in my query sequence that I did not put there?

You are seeing the result of automatic filtering of your query for low-complexity sequence that is performed to prevent artifactual hits. The filter substitutes any low-complexity sequence that it finds with the letter "N" in nucleotide sequence (e.g., "NNNNNNNNNNNNN") or the letter "X" in protein sequences (e.g., "XXXXXXXXX"). Low-complexity regions can result in high scores that reflect compositional bias rather than significant position-by-position alignment (Wootton amp; Federhen, 1996). Filter programs can eliminate these potentially confounding matches from the blast reports, leaving regions whose BLAST statistics reflect the specificity of their parities alignment. Queries searched with the blastn program are filtered with DUST. The other BLAST programs use SEG.

Q: How can I see low-similarity matches when there are many strong hits to my query sequence? Often, when the query is a member of a large sequence family, the summary hit list and the alignments returned only contain very high scoring hits. To look at low-similarity matches, you must increase the maximum number of results returned. On the BLAST Web pages, often it is sufficient to increase the size of the summary hit list and the number of alignments shown using the menus on the Advanced pages. However, it is possible to increase the lists even further using the Other Advanced Options box on the Advanced BLAST pages. For BLAST 2.0, "-v 2000", for example, will increase the number of descriptions returned in the summary hit list to 2000. The option "-b 2000" will similarly increase the number of alignments returned.

Q: I have heard that I will be penalized if I send a large number of sequences to the servers? .

The NCBI WWW BLAST server is a shared resource and it would be unfair for a few users to monoplize it. To prevent this, the server keeps track of how many queries are in the queue for each user and penalzies those users with many queries in the queue. This is done by calculating a 'Time of Execution' (TOE). If a user has only one query in the queue, then the TOE is set to the current time. As a user adds more queries to the queue, then the TOE is set to the current time, plus 60 seconds for every query in the queue. An example would be if a user sent in five requests one after the other without waiting for any to be worked on, then the TOE's for the requests would be:

1st request: current time

2nd request: current time + 60 seconds
3rd request: current time + 120 seconds
4th request: current time + 180 seconds
5th request: current time + 240 seconds

The BLAST server works through requests in the order of earliest to latest TOE. A query will be executed before it's TOE, if there are no other queries with an earlier TOE. Users with large numbers of queries are encouraged to use the BLAST servers at off-peaks hours, which are from 8 p.m. to 8 a.m. (EST).

NCBI

More Information

back to Education

BLAST
information
tutorial
guide

PSI-BLAST
tutorial

More Information
similarity searching
rules of thumb
glossary
reference list

Alignment
The process of lining up two or more sequences to achieve maximal levels of identity (and conservation, in the case of amino acid sequences) for the purpose of assessing the degree of similarity and the possibility of homology.

Algorithm
A fixed procedure embodied in a computer program.

Bioinformatics
The merger of biotechnology and information technology with the goal of revealing new insights and principles in biology.

Bit score
The value S' is derived from the raw alignment score S in which the statistical properties of the scoring system used have been taken into account. Because bit scores have been normalized with respect to the scoring system, they can be used to compare alignment scores from different searches.

BLAST
Basic Local Alignment Search Tool. (Altschul et al.) A sequence comparison algorithm optimized for speed used to search sequence databases for optimal local alignments to a query. The initial search is done for a word of length "W" that scores at least "T" when compared to the query using a substitution matrix. Word hits are then extended in either direction in an attempt to generate an alignment with a score exceeding the threshold of "S". The "T" parameter dictates the speed and sensitivity of the search. For additional details, see one of the BLAST tutorials (Query or BLAST) or the narrative guide to BLAST.

BLOSUM
Blocks Substitution Matrix. A substitution matrix in which scores for each position are derived from *observations* of the frequencies of substitutions in blocks of local alignments in related proteins. Each matrix is tailored to a particular evolutionary distance. In the BLOSUM62 matrix, for example, the alignment from which scores were derived was created using sequences sharing no more than 62% identity. Sequences more identical than 62% are represented by a single sequence in the alignment so as to avoid over-weighting closely related family members. (Henikoff and Henikoff)

Conservation
Changes at a specific position of an amino acid or (less commonly, DNA) sequence that preserve the physico-chemical properties of the original residue.

Domain
A discrete portion of a protein assumed to fold independently of

the rest of the protein and possessing its own function.

DUST

A program for filtering low complexity regions from nucleic acid sequences.

E value

Expectation value. The number of different alignents with scores equivalent to or better than S that are expected to occur in a database search by chance. The lower the E value, the more significant the score.

FASTA

The first widely used algorithm for database similarity searching. The program looks for optimal local alignments by scanning the sequence for small matches called "words". Initially, the scores of segments in which there are multiple word hits are calculated ("init1"). Later the scores of several segments may be summed to generate an "initn" score. An optimized alignment that includes gaps is shown in the output as "opt". The sensitivity and speed of the search are inversely related and controlled by the "k-tup" variable which specifies the size of a "word". (Pearson and Lipman)

Filtering

Also known as Masking. The process of hiding regions of (nucleic acid or amino acid) sequence having characteristics that frequently lead to spurious high scores. See SEG and DUST.

gap

A space introduced into an alignment to compensate for insertions and deletions in one sequence relative to another. To prevent the accumulation of too many gaps in an alignment, introduction of a gap causes the deduction of a fixed amount (the gap score) from the alignment score. Extension of the gap to encompass additional nucleotides or amino acid is also penalized in the scoring of an alignment.

Global Alignment

The alignment of two nucleic acid or protein sequences over their entire length.

H

H is the relative entropy of the target and background residue frequencies. (Karlin and Altschul, 1990). H can be thought of as a measure of the average information (in bits) available per position that distinguishes an alignment from chance. At high values of H, short alignments can be distinguished by chance, whereas at lower H values, a longer alignment may be necessary. (Altschul, 1991)

Homology

Similarity attributed to descent from a common ancestor.

HSP

High-scoring segment pair. Local alignments with no gaps that achieve one of the top alignment scores in a given search.

Identity
The extent to which two (nucleotide or amino acid) sequences are invariant.

K
A statistical parameter used in calculating BLAST scores that can be thought of as a natural scale for search space size. The value K is used in converting a raw score (S) to a bit score (S').

lambda
A statistical parameter used in calculating BLAST scores that can be thought of as a natural scale for scoring system. The value lambda is used in converting a raw score (S) to a bit score (S').

Local Alignment
The alignment of some portion of two nucleic acid or protein sequences

Low Complexity Region (LCR)
Regions of biased composition including homopolymeric runs, short-period repeats, and more subtle overrepresentation of one or a few residues. The SEG program is used to mask or filter LCRs in amino acid queries. The DUST program is used to mask or filter LCRs in nucleic acid queries.

Masking
Also known as Filtering. The removal of repeated or low complexity regions from a sequence in order to improve the sensitivity of sequence similarity searches performed with that sequence.

Motif
A short conserved region in a protein sequence. Motifs are frequently highly conserved parts of domains.

Multiple Sequence Alignment
An alignment of three or more sequences with gaps inserted in the sequences such that residues with common structural positions and/or ancestral residues are aligned in the same column. Clustal W is one of the most widely used multiple sequence alignment programs

Optimal Alignment
An alignment of two sequences with the highest possible score.

Orthologous
Homologous sequences in different species that arose from a common ancestral gene during speciation; may or may not be responsible for a similar function.

P value
The probability of an alignment occurring with the score in question or better. The p value is calculated by relating the observed alignment score, S, to the expected distribution of HSP scores from comparisons of random sequences of the same length and composition as the query to the database. The most highly significant P values will

be those close to 0. P values and E values are different ways of representing the significance of the alignment.

PAM

Percent Accepted Mutation. A unit introduced by Dayhoff et al. to quantify the amount of evolutionary change in a protein sequence. 1.0 PAM unit, is the amount of evolution which will change, on average, 1% of amino acids in a protein sequence. A PAM(x) substitution matrix is a look-up table in which scores for each amino acid substitution have been calculated based on the frequency of that substitution in closely related proteins that have experienced a certain amount (x) of evolutionary divergence.

Paralogous

Homologous sequences within a single species that arose by gene duplication.

Profile

A table that lists the frequencies of each amino acid in each position of protein sequence. Frequencies are calculated from multiple alignments of sequences containing a domain of interest. See also PSSM.

Proteomics

Systematic analysis of protein expression of normal and diseased tissues that involves the separation, identification and characterization of all of the proteins in an organism.

PSI-BLAST

Position-Specific Iterative BLAST. An iterative search using the BLAST algorithm. A profile is built after the initial search, which is then used in subsequent searches. The process may be repeated, if desired with new sequences found in each cycle used to refine the profile. Details can be found in this discussion of PSI-BLAST. (Altschul et al.)

PSSM

Position-specific scoring matrix; see profile. The PSSM gives the log-odds score for finding a particular matching amino acid in a target sequence.

Query

The input sequence (or other type of search term) with which all of the entries in a database are to be compared.

Raw Score

The score of an alignment, S, calculated as the sum of substitution and gap scores. Substitution scores are given by a look-up table (see PAM, BLOSUM). Gap scores are typically calculated as the sum of G, the gap opening penalty and L, the gap extension penalty. For a gap of length n, the gap cost would be G+Ln. The choice of gap costs, G and L is empirical, but it is customary to choose a high value for G (10-15)and a low value for L (1-2).

Similarity

RAFFINOSE SYNTHASE GENES AND THEIR USE

BACKGROUND OF THE INVENTION

FIELD OF INVENTION

The present invention relates to raffinose synthase genes and their use.

DISCLOSURE OF THE RELATED ART

Raffinose family oligosaccharides are derivatives of sucrose, which are represented by the general formula: o-α-D-galactopyranosyl-(1→6)$_n$-o-α-D-glucopyranosyl-(1→2)-β-D-fluctofuranoside, and they are called "raffinose" when n is 1, "stachyose" when n is 2, "verbascose" when n is 3, and "ajugose" when n is 4.

It has been known that raffinose family oligosaccharides have an effect of giving good conditions of enterobacterial flora, if present at an appropriate amount in food. Therefore, raffinose family oligosaccharides have already been used as a functional food material for addition to some kinds of food and utilized in the field of specific health food. On the other hand, raffinose family oligosaccharides are neither digested nor absorbed in mammals such as human, but are assimilated and decomposed by enterobacteria to generate gases and to cause meteorism and absorption disorder. Therefore, it has been desired to appropriately regulate the amount of raffinose family oligosaccharides in food and feed.

Raffinose family oligosaccharides are synthesized

by the raffinose family oligosaccharide biosynthesis system beginning with sucrose in many plants. This biosynthesis system normally involves a reaction for the sequential addition of galactosyl groups from galactinol through an α (1→6) bond to the hydroxyl group attached to the carbon atom at 6-position of the D-glucose residue in a sucrose molecule. Raffinose synthase is the enzyme concerned in the reaction for producing raffinose by allowing a D-galactosyl group derived from galactinol to form the α (1→6) bond with the hydroxyl group attached to the carbon atom at 6-position of the D-glucose residue in a sucrose molecule in the first step of this biosynthesis system. It has been suggested that this enzyme constitutes a rate-limiting step in the above synthesis system, and therefore this enzyme is quite important in the control of biosynthesis of raffinose family oligosaccharides.

Then, a method for controlling an expression level or activity of raffinose synthase in plants by utilizing a raffinose synthase gene is effective to control a biosynthesis system of raffinose family oligosaccharides in plants to increase or decrease the production of raffinose in plants. Thus, a raffinose synthase gene which can be used in such a method has been desired.

OBJECTS OF THE INVENTION

The main object of the present invention is to provide novel raffinose synthase genes from plants.

This object as well as other objects and advantage of the present invention will become apparent to those skilled in the art from the following description.

~~SUMMARY OF THE INVENTION~~

~~Under these circumstances, the present inventors have intensively studied and succeeded in isolating novel genes encoding raffinose synthase from various plants. Thus, the present invention has been completed.~~

~~That is, the present invention provides:~~

~~1. A raffinose synthase gene which comprises a nucleotide sequence hybridizable with a nucleotide sequence selected from the group consisting of:~~

~~(a) a nucleotide sequence encoding the amino acid sequence represented by SEQ ID NO: 1,~~

~~(b) the nucleotide sequence represented by SEQ ID NO: 2,~~

~~(c) a nucleotide sequence encoding the amino acid sequence of represented by SEQ ID NO: 3,~~

~~(d) the nucleotide sequence represented by the 236th to 2584th nucleotides in the nucleotide sequence represented by SEQ ID NO: 4,~~

~~(e) a nucleotide sequence encoding the amino acid sequence represented by SEQ ID NO: 5,~~

~~(f) the nucleotide sequence represented by the 134th to 2467th nucleotides in the nucleotide sequence represented~~

~~by SEQ ID NO: 6,~~

~~(g) a nucleotide sequence encoding the amino acid sequence represented by SEQ ID NO: 7, and~~

~~(h) the nucleotide sequence represented by the 1st to 1719th nucleotides in the nucleotide sequence represented by SEQ ID NO: 8,~~

~~under stringent conditions, and encoding a protein being capable of binding D-galactosyl group through α (1→6) bond to the hydroxyl group attached to the carbon atom at 6-position of the D-glucose residue in a sucrose molecule to form raffinose.~~

~~2. A raffinose synthase gene comprising a nucleotide sequence encoding the amino acid sequence represented by SEQ ID NO: 1.~~

~~3. A raffinose synthase gene comprising the nucleotide sequence represented by SEQ ID NO: 2.~~

~~4. A raffinose synthase gene comprising a nucleotide sequence encoding the amino acid sequence represented by SEQ ID NO: 3.~~

~~5. A raffinose synthase gene comprising the nucleotide sequence represented by the 236th to 2584th nucleotides in the nucleotide sequence represented by SEQ ID NO: 4.~~

~~6. A raffinose synthase gene comprising a nucleotide sequence encoding the amino acid sequence represented by SEQ ID NO: 5.~~

~~7. A raffinose synthase gene comprising the~~

~~nucleotide sequence represented by the 134th to 2467th nucleotides in the nucleotide sequence represented by SEQ ID NO: 6.~~

~~8. A raffinose synthase gene comprising a nucleotide sequence encoding the amino acid sequence represented by SEQ ID NO: 7.~~

~~9. A raffinose synthase gene comprising the nucleotide sequence represented by the 1st to 1719th nucleotides in the nucleotide sequence represented by SEQ ID NO: 8.~~

~~10. A raffinose synthase gene comprising the nucleotide sequence represented by SEQ ID NO: 4, SEQ ID NO: 6 or SEQ ID NO: 8.~~

~~11. A nucleic acid comprising a partial nucleotide sequence of the raffinose synthase gene of any one of the above 1 to 10.~~

~~12. A method for detecting a nucleic acid containing a raffinose synthase gene which comprises detecting said nucleic acid by hybridization using the labeled nucleic acid of the above 11 as a probe.~~

~~13. A method for amplifying a nucleic acid containing a raffinose synthase gene which comprises amplifying said nucleic acid by polymerase chain reaction (PCR) using the nucleic acid of the above 11 as a primer.~~

~~14. A method for obtaining a raffinose synthase gene which comprises the steps of:~~

~~detecting a nucleic acid containing said raffinose~~

~~synthase gene by hybridization using the labeled nucleic acid of the above 11 as a probe, and~~

~~recovering the detected nucleic acid.~~

~~15. A method for obtaining a raffinose synthase gene which comprises the steps of:~~

~~amplifying a nucleic acid containing said raffinose synthase gene by PCR using the nucleic acid of the above 11 as a primer, and~~

~~recovering the amplified nucleic acid.~~

~~16. A nucleic acid comprising a nucleic acid containing the raffinose synthase gene of any one of the above 1 to 10 which is joined to a nucleic acid exhibiting promoter activity in a host cell.~~

~~17. A vector comprising the raffinose synthase gene of any one of the above 1 to 10.~~

~~18. A transformant, wherein the raffinose synthase gene of any one of the above 1 to 10 is introduced into a host cell.~~

~~19. A transformant, wherein the nucleic acid of the above 16 is introduced into a host cell.~~

~~20. A transformant, wherein the vector of the above 17 is introduced into a host cell.~~

~~21. The transformant of any one of the above 18 to 20, wherein the host is a microorganism.~~

~~22. The transformant of any one of the above 18 to 20, wherein the host is a plant.~~

~~23. A method for producing a raffinose synthase which comprises the steps of:~~

~~culturing or growing the transformant of any one of the above 18 to 22 to produce the raffinose synthase, and~~

~~collecting the raffinose synthase.~~

~~24. A raffinose synthase comprising the amino acid sequence represented by SEQ ID NO: 1.~~

~~25. A raffinose synthase comprising the amino acid sequence represented by SEQ ID NO: 3.~~

~~26. A raffinose synthase comprising the amino acid sequence represented by SEQ ID NO: 5.~~

~~27. A raffinose synthase comprising the amino acid sequence represented by SEQ ID NO: 7.~~

~~The term "nucleic acid" used herein means an oligomer compound or a high molecular compound which is generally called "DNA" or "RNA".~~

DETAILED DESCRIPTION OF THE INVENTION

The gene engineering techniques described below can be carried out, for example, according to methods described in "Molecular Cloning: A Laboratory Manual 2nd edition" (1989), Cold Spring Harbor Laboratory Press, ISBN 0-87969-309-6; "Current Protocols In Molecular Biology" (1987), John Wiley & Sons, Inc. ISBN 0-471-50338-X; "Current Protocols In Protein Science" (1995), John Wiley & Sons, Inc. ISBN 0-471-11184-8.

The genes of the present invention can be obtained

from soybean, plants belonging to the families Chenopdiaceae such as beet, etc. and Cruciferae such as mustard, rapeseed, etc. Specific examples of the genes of the present invention include those comprising a nucleotide sequence encoding the amino acid sequence represented by SEQ ID NO: 1, the nucleotide sequence represented by SEQ ID NO: 2, a nucleotide sequence encoding the amino acid sequence represented by SEQ ID NO: 3, the nucleotide sequence represented by SEQ ID NO: 4 or by the 236th to 2584th nucleotides in the nucleotide sequence represented by SEQ ID NO: 4, a nucleotide sequence encoding the amino acid sequence represented by SEQ ID NO: 5, the nucleotide sequence represented by SEQ ID NO: 6 or by the 134th to 2467th nucleotides in the nucleotide sequence represented by SEQ ID NO: 6, a nucleotide sequence encoding the amino acid sequence represented by SEQ ID NO: 7, the nucleotide sequence represented by SEQ ID NO: 8 or by the 1st to 1719th nucleotides in the nucleotide sequence represented by SEQ ID NO: 8, and the like.

The genes of the present invention can be obtained, for example, by the following method.

That is, the genes of the present invention derived from soybean can be obtained, for example, by the following method.

For example, the gene can be obtained by a hybridization method using a nucleic acid having the nucleotide sequence represented by SEQ ID NO: 2 or its partial nucleotide sequence as a probe to detect a nucleic acid fragment which hybridizes

to the probe in DNAs derived from soybean, followed by isolating the detected nucleic acid.

In this method, first, a nucleic acid to be used as the probe is prepared. As such a nucleic acid, for example, there is a nucleic acid composed of an oligonucleotide chemically synthesized by a conventional method on the basis of the nucleotide sequence of SEQ ID NO: 2. Specific example thereof includes a nucleic acid having the 800th to the 899th nucleotides in the nucleotide sequence represented by SEQ ID NO: 2.

Alternatively, the gene of the present invention derived from soybean can be obtained by the following method.

For example, tissue of soybean (*Glycine max*) is frozen in liquid nitrogen and ground physically with a mortar or other means into finely divided tissue debris powder. From the tissue debris powder, RNA is extracted by a conventional method. A commercially available RNA extraction kit can be utilized in the extraction. RNA is recovered from thus-obtained RNA extract by ethanol precipitation. Poly-A tailed RNA is fractionated from thus-recovered RNA by a conventional method. A commercially available oligo-dT column can be utilized in this fractionation. cDNA is synthesized from the poly-A tailed RNA thus obtained by a conventional method. The synthesis can be carried out by using a commercially available cDNA synthesis kit. DNA is amplified by PCR using the above-obtained cDNA as the template and primers designed and synthesized on the basis of the nucleotide sequence of SEQ ID NO: 2. More specifically, as the primers,

for example, there are primers 11 (SEQ ID NO: 9) and 12 (SEQ ID NO: 10) shown in List 1 hereinafter. When PCR is carried out by using these primers and as the template cDNA derived from soybean, the genes of the present invention derived from soybean, e.g., the "raffinose synthase gene having a nucleotide sequence encoding the amino acid sequence of SEQ ID NO: 1" and the "raffinose synthase gene having the nucleotide sequence of SEQ ID No: 2" can be obtained.

The amplified DNA can be cloned according to a conventional method, for example, described in "Molecular Cloning: A Laboratory Manual 2nd edition" (1989), Cold Spring Harbor Laboratory Press; or "Current Protocols In Molecular Biology" (1987), John Wiley & Sons, Inc. ISBN 0-471-50338-X. Alternatively, cloning can be carried out, for example, by using a commercially available cloning kit such as TA cloning kit (Invitrogen) and a commercially available plasmid vector such as pBluescript II (Stratagene). The nucleotide sequence of the DNA clone can be determined by dideoxy terminating method such as that described by F. Sanger, S. Nicklen, A.R. Coulson, Proceedings of National Academy of Science U.S.A. (1977), 74, pp. 5463-5467. For example, preferably, a commercially available kit such as ABI PRISM Dye Terminator Cycle Sequencing Ready Reaction Kit manufactured by Perkin-Elmer can be used.

List 1

Primer 11 (SEQ ID NO: 9): ccaatctgat catgcttgtg ccgaa

25mer

Primer 12 (SEQ ID NO: 10): ggaacaaagt tatgcactat tatttaaggt 30mer

The genes of the present invention derived from a Chenopdiaceae plant such as beet can be obtained by the following method.

For example, tissue of a Chenopdiaceae plant such as beet (*Beta vulgaris*) is frozen in liquid nitrogen and ground physically with a mortar or other means into finely divided tissue debris powder. From this tissue debris powder, RNA is extracted by a conventional method. A commercially available RNA extraction kit can be utilized in the extraction. RNA is recovered from the thus-obtained RNA extract by ethanol precipitation. From the recovered RNA, poly-A tailed RNA is fractionated by a conventional method. A commercially available oligo-dT column can be utilized in this fractionation. cDNA is synthesized from the poly-A tailed RNA thus obtained by a conventional method. The synthesis can be carried out by utilizing a commercially available cDNA synthesis kit. DNA is amplified by PCR using the above-obtained cDNA as the template and primers designed and chemically synthesized on the basis of the nucleotide sequence of SEQ ID NO: 4. More specifically, as the primers, for example, there are primers 21 (SEQ ID NO: 11) and 22 (SEQ ID NO: 12) shown in List 2 hereinafter. When PCR is carried out by using these primers and as the template cDNA derived from beat, the genes of the present invention derived from beet, e.g., the "raffinose synthase gene having a nucleotide

sequence encoding the amino acid sequence of SEQ ID NO: 3," and the "raffinose synthase gene having a nucleotide sequence of SEQ ID No: 4" can be obtained. According to a particular purpose, the PCR primers can also be designed and synthesized on the basis of the nucleotide sequence of SEQ ID NO: 4. For example, in order to amplify the "raffinose synthase gene having the nucleotide sequence represented by the 236th to the 2584th nucleotides in the nucleotide sequence represented by SEQ ID NO: 4", preferably, oligonucleotides having the nucleotide sequences represented by primers 23 (SEQ ID NO: 13) and 24 (SEQ ID NO: 14) in List 2 below are synthesized and used as the primers.

The amplified DNA can be cloned according to a conventional method, for example, described in "Molecular Cloning: A Laboratory Manual 2nd edition" (1989), Cold Spring Harbor Laboratory Press; or "Current Protocols In Molecular Biology" (1987), John Wiley & Sons, Inc. ISBN 0-471-50338-X. Alternatively, cloning can be carried out by using a commercially available cloning kit such as TA cloning kit (Invitrogen) and a commercially available plasmid vector such as pBluescript II (Stratagene). The nucleotide sequence of the DNA clone can be determined, for example, by dideoxy terminating method such as that described by F. Sanger, S. Nicklen, A.R. Coulson, Proceedings of National Academy of Science U.S.A. (1977), 74, pp. 5463-5467. For example, preferably, a commercially available kit such as ABI PRISM Dye Terminator Cycle Sequencing Ready Reaction Kit manufactured by Perkin-Elmer can be used.

List 2

Primer 21 (SEQ ID NO: 11): ctaccaaatt ccacaactta aagttca 27mer

Primer 22 (SEQ ID NO: 12): ggaataataa gcttcacaca tactgtactc tc 32mer

Primer 23 (SEQ ID NO: 13): atggctccaa gctttagcaa ggaaaattcc 30mer

Primer 24 (SEQ ID NO: 14): tcaaaataag tactcaacag tggtaaaacc 30mer

The genes of the present invention derived from Cruciferae plants such as mustard (*Brassica juncea*) and rapeseed (*Brassica napus*) can be obtained by the following method.

For example, tissue of a Cruciferae plant such as mustard or rapeseed is frozen in liquid nitrogen and ground physically with a mortar or other means into finely divided tissue debris powder. From the tissue debris powder, RNA is extracted by a conventional method. A commercially available RNA extraction kit can be utilized in the extraction. The RNA is recovered from thus-obtained RNA extract by ethanol precipitation. Poly-A tailed RNA is fractionated from the RNA thus recovered by a conventional method. A commercially available oligo-dT column can be utilized in the fractionation. cDNA is synthesized from the poly-A tailed RNA thus obtained by a conventional method. The synthesis can be carried out by using a commercially available cDNA synthesis kit. DNA are amplified by PCR using the above-obtained cDNA as a template

and primers designed and chemically synthesized on the basis of the nucleotide sequence of SEQ ID NO: 6. For example, when PCR is carried out by using cDNA derived from mustard (*Brassica juncea*) as the template and primers 33 (SEQ ID NO: 17) and 34 (SEQ ID NO: 18) shown in List 3 hereinafter, the genes from Cruciferae plants of the present invention, e.g., the "raffinose synthase gene having a nucleotide sequence encoding the amino acid sequence of SEQ ID NO: 5," and the "raffinose synthase gene having the nucleotide sequence represented by the 1st to 2654th nucleotides in the nucleotide sequence represented by SEQ ID NO: 6" can be obtained. According to a particular purpose, the PCR primers can also be designed and synthesized on the basis of the nucleotide sequence of SEQ ID NO: 6. For example, in order to amplify DNA encoding the open reading frame region of the "raffinose synthase gene having a nucleotide sequence encoding a protein having the amino acid sequence of SEQ ID NO: 5", and the "raffinose synthase gene having the nucleotide sequence represented by the 134th to the 2467th nucleotides of SEQ ID NO: 6", preferably, oligonucleotides having the nucleotide sequences represented by primers 35 (SEQ ID NO: 19) and 36 (SEQ ID NO: 20) in List 3 are synthesized and used as the primers.

The amplified DNA can be cloned according to a conventional method, for example, described in "Molecular Cloning: A Laboratory Manual 2nd edition" (1989), Cold Spring Harbor Laboratory Press; or "Current Protocols In Molecular

Biology" (1987), John Wiley & Sons, Inc. ISBN 0-471-50338-X. Alternatively, cloning can be carried out, for example, by using a commercially available TA cloning kit (Invitrogen) or a commercially available plasmid vector such as pBluescript II (Stratagene). The nucleotide sequence of the DNA clone can be determined by dideoxy terminating method such as that described by F. Sanger, S. Nicklen, A.R. Coulson, Proceedings of National Academy of Science U.S.A. (1977), 74, pp. 5463-5467. For example, preferably, the commercially available ABI PRISM Dye Terminator Cycle Sequencing Ready Reaction Kit of Perkin-Elmer can be used.

List 3

Primer 31 (SEQ ID NO: 15): ttggaagaga agacgccgcc gggaatcgtc 30mer

Primer 32 (SEQ ID NO: 16): ttaagccccg gcgagagctc tggccggaca 30mer

Primer 33 (SEQ ID NO: 17): accaatccaa aatctcatca aataatcgca 30mer

Primer 34 (SEQ ID NO: 18): aaataatagg ggcagtacaa attacaccac 30mer

Primer 35 (SEQ ID NO: 19): atggctccac cgagcgtaat taaatccga 29mer

Primer 36 (SEQ ID NO: 20): ctaaaactca tacttaatag aagacaaacc 30mer

Then, a nucleic acid having a partial nucleotide sequence of the gene of the present invention (hereinafter referred to as "the gene fragment") which is obtained by the

above-described method is labeled and then used as a probe in a hybridization method. The probe can be hybridized to, for example, DNA derived from soybean, a Chenopdiaceae plant or a Cruciferae plant to detect a nucleic acid having the probe specifically bound thereto, thereby detecting a nucleic acid having the raffinose synthase gene.

As the DNA derived from soybean, a Chenopdiaceae plant such as beet or a Cruciferae plant such as mustard or rapeseed, for example, a cDNA library or a genomic DNA library of these plants can be used. The gene library may also be a commercially available gene library as such or a library constructed according to a conventional library construction method, for example, described in "Molecular Cloning: A Laboratory Manual 2nd edition" (1989), Cold Spring Harbor Laboratory Press; "Current Protocols In Molecular Biology" (1987), John Wiley & Sons, Inc. ISBN 0-471-50338-X.

As the hybridization method, for example, plaque hybridization or colony hybridization can be employed, and they are selected depending upon the kind of vector used in the construction of a library. More specifically, when the library to be used is constructed with a phage vector, a suitable host microorganism is mixed with the phage of the library under infectious conditions to obtain transformants. The transformant is further mixed with a soft agar medium, and the mixture is plated on an agar medium. Thereafter, the mixture is cultured at 37°C until a plaque of an appropriate size appears.

When the library to be used is constructed with a plasmid vector, the plasmid is introduced into a suitable host microorganism to form transformants. The transformant obtained is diluted to a suitable concentration and the dilution is plated on an agar medium, after which it is cultured at 37°C until a colony of an appropriate size appears. In either case of the above libraries, a membrane filter is placed on the surface of the agar medium after the above cultivation, so that the phage or transformant is transferred to the membrane. This membrane is denatured with an alkali, followed by neutralization, and for example, when a nylon membrane is used, the membrane is irradiated with ultraviolet light, so that DNA of the phage or transformant is fixed on the membrane. This membrane is then subjected to a hybridization method wherein the gene fragment which has a partial nucleotide sequence of the gene of the present invention and labeled by a conventional method (hereinafter referred to as "the labeled gene fragment") is used as a probe. For this method, reference may be made, for example, to D.M. Glover ed., "DNA cloning, a practical approach" IRL PRESS (1985), ISBN 0-947946-18-7. There are various reagents and temperature conditions to be used in the hybridization. For example, in general, prehybridization is carried out by immersion of the membrane in a prehybridization solution [6 x SSC (0.9 M NaCl, 0.09 M citric acid), 0.1 to 1 (w/v)% SDS, 100 μg/ml denatured salmon sperm DNA] and incubation at 65°C for 1 hour. Then, hybridization is carried out by addition and mixing of the labeled

gene fragment thereto and incubating the membrane at 42 to 68°C for 4 to 16 hours.

In the present invention, the "stringent conditions" are those wherein incubation is carried out, for example, at 65 to 68°C in the above hybridization.

After hybridization, the membrane is taken out and is washed with 2 x SSC containing 0.1 to 1 (w/v)% SDS, further rinsed with 0.2 x SSC containing 0.1 to 1 (w/v)% SDS, and then dried. The membrane is analyzed, for example, by autoradiography or other techniques to detect the position of the probe on the membrane, thereby detecting the position on the membrane of a nucleic acid having a nucleotide sequence homologous to that of the probe used. The clone corresponding to the position of the nucleic acid thus detected on the membrane is identified on the original agar medium and the positive clone is selected so that the clone having the nucleic acid can be isolated. The same procedures of detection are repeated to purify the clone having the nucleic acid.

Alternatively, a commercially available kit such as GENE TRAPPER cDNA Positive Section System kit (GibcoBRL) can be used. In this method, first, a single-stranded DNA library is hybridized with the biotynylated gene fragment (i.e., probe), followed by adding streptoavidin-bound magnet beads and mixing. From the mixture, the streptoavidin-bound magnetic beads are collected with a magnet, so that single-stranded DNA having a nucleotide sequence homologous to that of the probe used,

which has been bound to these beads through the gene fragment, biotin and streptoavidin, is collected and detected. The single-stranded DNA collected can be converted into a double-strand form by reaction with a suitable DNA polymerase using a suitable oligonucleotide as a primer.

As described above, a nucleic acid containing raffinose synthase gene can obtained by detecting a nucleic acid hybridizable to the gene fragment in DNAs of a gene library derived from soybean, a Chenopdiaceae plant or a Cruciferae plant, purifying a clone having the nucleic acid and isolating phage or plasmid DNA from the clone. By preparing the restriction map or determining the nucleotide sequence of the nucleic acid thus obtained according to a conventional method, the nucleic acid containing the gene of the present invention can be confirmed.

For example, the gene of the present invention from a Chenopdiaceae plant can be confirmed by the following point:

The amino acid encoded by the nucleotide sequence thus determined has 75% or more homology to the amino acid sequence represented by the 103rd to 208th amino acids in the amino acid sequence of SEQ ID NO: 3;

80% or more homology to the amino acid sequence represented by the 255th to 271st amino acids in the amino acid sequence of SEQ ID NO:3;

70% or more homology to the amino acid sequence represented by the 289th to 326th amino acids in the amino acid sequence of SEQ ID NO: 3; or

70% or more homology to the amino acid sequence represented by the 610th to 696th amino acids in the amino acid sequence of SEQ ID NO: 3.

The gene of the present invention from a Cruciferae plant can be confirmed, for example, by the following point:

The amino acid sequence encoded by the nucleotide sequence determined has 75% or more homology to the amino acid sequence represented by the 111th to 213th amino acids in the amino acid sequence of SEQ ID NO: 5;

80% or more homology to the amino acid sequence represented by the 260th to 275th amino acids in the amino acid sequence of SEQ ID NO: 5;

70% or more homology to the amino acid sequence represented by the 293rd to 325th amino acids in the amino acid sequence of SEQ ID NO: 5; or

70% or more homology to the amino acid sequence represented by the 609th to 695th amino acids in the amino acid sequence of SEQ ID NO: 5.

The "homology" used herein means the proportion of the number of amino acids in a region, which are identical to those in a different region to be compared, to the number of the entire amino acids in the former region, upon comparing regions having similarity in two amino acid sequences. In this respect, it is preferred that the region having similarity contains more amino acids. Such homology of amino acid sequences can be evaluated by using a commercially available gene analysis

software such as GENETIX (Software Kaihatu K.K.).

Further, according to the same manner as described above, a nucleic acid containing raffinose synthase gene can be detected by hybridization to DNA from the desired organism using the gene fragment as a probe to detect a nucleic acid to which the probe specifically binds (hereinafter referred to as the detection method of the present invention). The gene fragment used herein can be chemically synthesized according to a conventional method on the basis of the nucleotide sequence represented by SEQ ID NO: 2, 4, 6 or 8. Alternatively, it can be prepared by PCR using as primers oligonucleotides chemically synthesized according to a conventional method on the basis of the nucleotide sequence represented by SEQ ID NO: 2, 4, 6 or 8.

The gene fragment may be a part of the non-translated region of the raffinose synthase gene as well as the open reading frame thereof. For example, an oligonucleotide having the same nucleotide sequence as a part of that of 5'-upstream side such as the 1st to 235th nucleotides in the nucleotide sequence of SEQ ID NO: 4, the 1st to 133rd nucleotides in the nucleotide sequence of SEQ ID NO: 6 and the like, or a part of that of 3'-downstream side such as the 2588th to 2675th nucleotides in the nucleotide sequence of SEQ ID NO: 4, the 2468th to 2676th nucleotides in the nucleotide sequence of SEQ ID NO: 6 and the like.

When PCR is carried out by using the gene fragment

as primers, it is possible to amplify a nucleic acid containing raffinose synthase gene from DNA derived from the desired organism (hereinafter referred to as the amplification method of the present invention).

More specifically, for example, oligonucleotides having the nucleotide sequences of the gene fragment are designed and chemically synthesized according to a conventional method. In general, it is preferred that the number of nucleotides is more from a viewpoint that the specificity of annealing is ensured. It is, however, also preferred that the number of nucleotides is not so many from viewpoints that the primers themselves are liable to have a higher structure giving possible deterioration of the annealing efficiency and that complicated procedures are required in the purification after the synthesis. Normally, oligonucleotides composed of 15 to 50 bases are preferred. In this respect, based on the codon table showing the correspondence of amino acids encoded by codons, a mixture of primers can also be synthesized by using a mixture of plural bases so that a residue at a specified position in a primer is changed to different bases according to the variation of codons which can encode one certain amino acid. Alternatively, for example, a base such as inosine which can form a base pair with plural bases can be used instead of the above mixture of plural bases.

Coding Table

Phe: UUU, UUC Ser: UCU, UCC, UCA, UCG, AGU, AGC

Tyr: UAU, UAC Cys: UGU, UGC

Stop: UAA, UAG, UGA Trp: UGG

Leu: UUA, UUG, CUU, CUC, CUA, CUG Pro: CCU, CCC, CCA, CCG His: CAU, CAC Gln: CAA, CAG Arg: CGU, CGC, CGA, CGG, AGA, AGG

Ile: AUU, AUC, AUA Thr: ACU, ACC, ACA, ACG

Asn: AAU, AAC Lys: AAA, AAG

Met: AUG

Val: GUU, GUC, GUA, GUG Ala: GCU, GCC, GCA, GCG

Asp: GAU, GAC Gly: GGU, GGC, GGA, GGG

Glu: GAA, GAG

In the above codon table, each codon is shown as the nucleotide sequence in mRNA and its light hand is the 5'-terminus. U represents uracil base in RNA and corresponds to thymine base in DNA.

An oligonucleotide having the same nucleotide sequence as the coding strand of the double-stranded DNA of the gene of the present invention is called a "sense primer" and that having a nucleotide sequence complementary to the coding strand is called an "antisense primer".

A sense primer having the same nucleotide sequence as that of 5'-upstream side in the coding strand of the gene of the present invention, and an antisense primer having a nucleotide sequence complementary to the nucleotide sequence on the 3'-downstream side in this coding strand are used in combination for PCR reaction, for example, with a gene library,

genomic DNA or cDNA as the template to amplify DNA. As the gene library to be used, for example, there are a cDNA library and a genomic library derived from soybean, a Chenopdiaceae plant such as beet or a Cruciferae plant such as mustard or rapeseed, etc. The gene library may also be a library constructed according to a conventional library construction method, for example, described in "Molecular Cloning: A Laboratory Manual 2nd edition" (1989), Cold Spring Harbor Laboratory Press; "Current Protocols In Molecular Biology" (1987), John Wiley & Sons, Inc. ISBN 0-471-50338-X, or a commercially available gene library as such. As the genomic DNA or cDNA, for example, there are those prepared from soybean, a Chenopdiaceae plant such as beet or a Cruciferae plant such as mustard or rapeseed, etc. For example, PCR is carried out by using the primers 31 (SEQ ID NO: 15) and 32 (SEQ ID NO: 16) in the above List 3 and as the template cDNA derived from mustard to amplify DNA having the nucleotide sequence represented by the 749th to 1215th nucleotides in the nucleotide sequence of SEQ ID NO: 6. Further, PCR is carried out by using the primers and as the template cDNA derived from rapeseed to amplify DNA having the nucleotide sequence represented by the 1st to 467th nucleotides in the nucleotide sequence of SEQ ID NO: 8. The nucleic acid thus amplified can be confirmed by conventional electrophoresis. The nucleic acid can be cloned according a conventional method such as that described in "Molecular Cloning: A Laboratory Manual 2nd edition" (1989), Cold Spring Harbor laboratory Press or

"Current Protocols in Molecular Biology" (1987), John Wiley & Sons, Inc., ISBN 0-471-50338-X. For the nucleic acid, its restriction map is prepared or its nucleotide sequence is determined by a conventional method, so that the nucleic acid containing raffinose synthase gene or a part thereof can be identified. When the nucleic acid contains a part of raffinose synthase, PCR can be carried out on the basis of its nucleotide sequence to amplify the nucleic acid containing the 5'-upstream side nucleotide sequence or the 3'-downstream side nucleotide sequence. That is, based on the nucleotide sequence of the above-obtained nucleic acid, an antisense primer is designed and synthesized for amplification of the 5'-upstream side part, and a sense primer is designed and synthesized for amplification of the 3'-downstream side part. The nucleotide sequence of the 5'-upstream side part or 3'-downsream side part of the nucleotide sequence already obtained can be determined by the RACE method using these primers and a commercially available kit such as Marathon Kit of Clontech. The full length raffinose synthase gene can be obtained by synthesizing new primers based on both terminal sequences in the nucleotide sequence thus determined and carrying out PCR again.

The above detection method of the present invention can also be used in the analysis of genotypes of a plant such as soybean, a Chenopdiaceae plant or a Cruciferae plant, etc. More specifically, for example, a genomic DNA derived from soybean, a Chenopdiaceae plant or a Cruciferae plant is prepared

according to a conventional method, for example, described in "Cloning and Sequence (Plant Biotechnology Experiment Manual)" complied under the supervision of Itaru Watanabe, edited by Masahiro Sugiura, published by Noson Bunka-sha, Tokyo (1989). The genomic DNA is digested with at least several kinds of restriction enzymes, followed by electrophoresis. The electrophoresed DNA is blotted on a filter according to a conventional method. This filter is subjected to hybridization with a probe prepared from DNA having the gene fragment by a conventional method, and DNA to which the probe hybridizes is detected. The DNAs detected are compared in length between different varieties of a specified plant species. The differences in length make possible the analysis of differences in phenotypic characteristics accompanied with the expression of raffinose family oligosaccharides between these varieties. Furthermore, when the DNAs detected by the above method are compared in length between the gene recombinant plant and the non-gene recombinant plant of the same variety, the former plant can be distinguished from the latter plant by the detection of hybridizing bands greater in number or higher in concentration for the former plant than for the latter plant. This method can be carried out according to the RFLP (restriction fragment length polymorphism) method, for example, described in "Plant PCR Experiment Protocols" complied under the supervision of Ko Shimamoto and Takuji Sasaki, published by Shujun-sha, Tokyo (1995), ISBN 4-87962-144-7, pp. 90-94.

Further, the amplification method of the present invention can be used for an analysis of genes of soybean, a Chenopdiaceae plant or a Cruciferae plant, etc. More specifically, for example, the amplification method of the present invention is carried out by using plant genomic DNA prepared from soybean, a Chenopdiaceae plant or a Cruciferae plant to amplify DNA. The amplified DNA is mixed with a formaldehyde solution, followed by heat denaturing at 85°C for 5 minutes and then quickly cooling on ice. A sample thereof is subjected to electrophoresis on, for example, 6 (w/v)% polyacrylamide gel containing 0 (v/v)% or 10 (v/v)% of glycerol. For this electrophoresis, a commercially available electrophoresis apparatus such as that for SSCP (Single Strand Conformation Polymorhism) can be used and the electrophoresis can be carried out with maintaining the gel at a constant temperature, for example, at 5°C, 25°C, 37°C, etc. From the electrophoresed gel, DNA is detected, for example, by a method such as silver staining method with a commercially available reagent. From the differences of behavior between the varieties in the electrophoresis of the DNA detected, a mutation in the raffinose synthase gene is detected, and an analysis is carried out for differences caused by the mutation in phenotypic characteristics accompanied with the expression of raffinose family oligosaccharides. This method can be carried out according to the SSCP method, for example, described in "Plant PCR Experiment Protocols" complied under the supervision of

Ko Shimamoto and Takuji Sasaki, published by Shujun-sha, Tokyo (1995), ISBN 4-87962-144-7, pp. 141-146.

The analysis of the plant gene from soybean, a Chenopdiaceae plant or a Cruciferae plant by the above detection method or amplification method of the present invention can be used not only for the analysis of differences in phenotypic characteristics accompanied with the expression of raffinose family oligosaccharides, but also, for example, for the selection of clones having the desired characters upon production of a novel variety of soybean, a Chenopdiaceae plant or a Cruciferae plant. Further, it can also be used for identification of a clone thus produced and having the characters derived from a recombinant plant upon producing a plant variety using the recombinant plant.

For expression of the gene of the present invention in cells of a host, preferably, a nucleic acid comprising a nucleic acid fragment which contains the gene of the present invention, and a nucleic acid fragment which has a promoter activity in the host cells and joined to the former nucleic acid fragment (hereinafter referred to as the expression nucleic acid of the present invention) can be used.

The nucleic acid fragment having promoter activity in the expression nucleic acid of the present invention is not limited to a specific one, so long as it is functionable in a host to be transformed. For example, there are synthetic promoters functionable in *Escherichia coli*, such as *E. coli*

lactose operon promoter, *E. coli* tryptophan operon promoter and tac promoter, etc.; yeast alcohol dehydrogenase gene (ADH) promoter, adenovirus major late (Ad.ML) promoter, SV40 early promoter, baculovirus promoter and the like. When the host is a plant, the promoter may include, for example, T-DNA derived constitutive promoters such as nopaline synthase gene (NOS) promoter, octopine synthase gene (OCS) promoter, etc.; plant virus-derived promoters such as cauliflower mosaic virus (CaMV)-derived 19S and 35S promoters; inducible promoters such as phenylalanine ammonia-lyase (PAL) gene promoter, chalcone synthase (CHS) gene promoter, pathogenesis-related protein (PR) gene promoter, etc. Furthermore, vector pSUM-GY1 (see JP-A 06-189777/1994) can also be used, which has a promoter giving specific expression in a specified plant tissue, e.g., a promoter of soybean-derived seed storage protein glycinin gene (JP-A 6-189777).

Furthermore, a nucleic acid fragment having a terminator activity can be joined to the expression nucleic acid of the present invention. In this case, it is generally preferred that the expression nucleic acid of the present invention is constructed so that the nucleic acid fragment having a terminator activity is positioned downstream the raffinose synthase gene. The terminator to be used is not particularly limited, so long as it is functionable in cells of a host to be transformed. For example, when the host is a plant, there are T-DNA derived constitutive terminators such as nopaline

synthase gene (NOS) terminator, etc.; plant derived terminators such as terminators of allium virus GV1 or GV2, and the like.

The expression nucleic acid of the present invention can be introduced into a host cell according to a conventional gene engineering technique to obtain a transformant. If necessary, the expression nucleic acid of the present invention can be inserted into a vector having a suitable marker depending upon a particular transformation technique for introduction of the nucleic acid into a host cell.

A vector into which the expression nucleic acid of the present invention is inserted can be introduced into a microorganism according to a conventional method, for example, described in "Molecular Cloning: A Laboratory Manual 2nd edition" (1989), Cold Spring Harbor laboratory Press or "Current Protocols in Molecular Biology" (1987), John Wiley & Sons, Inc., ISBN 0-471-50338-X. The microorganism transformed with the vector can be selected on the basis of a selection marker such as antibiotic resistance, auxotrophy or the like. In case that the gene of the present invention is joined to the downstream of an inducible promoter (e.g., tac promoter) in the translatable form in the selected microorganism (e.g., *E. coli* transformant), a translated product of the gene of the present invention can be expressed under conventional culture and inducible conditions and can be recover as a peptide or a protein.

The raffinose synthase activity of the translated product of the gene of the present invention thus prepared can

be measured by, for example, a method described in L. Lehle and W. Tanner, Eur. J. Biochem., 38, 103-110 (1973) to identify the translated product having the "capability of binding D-galactosyl group through α (1→6) bond to the hydroxyl group attached to the carbon atom at 6-position of the D-glucose residue in sucrose molecule". More specifically, for example, the gene of the present invention is cloned in pGEX-4T3 (Pharmacia) to obtain a plasmid containing the expression nucleic acid of the present invention. The resultant plasmid is introduced into, for example, *E. coli* HB101 strain to obtain a transformant. The resultant transformant is culture overnight and 1 ml of the culture is inoculated into 100 ml of LB culture medium. It is incubated at 37°C for about 3 hours and IPTG (isopropylthio-β-D-galactoside) is added at a final concentration of 1 mM, followed by further incubation for 5 hours. Cells are recovered from the culture broth by centrifugation and are suspended by addition of 10 times of the cell weight of 100 mM Tris-HCl (pH 7.4), 1 mM EDTA, 5 mM DTT (dithiothreitol), 1 mM PMSF (phenylmethylsulfonyl fluoride) and 1 mM benzamide. The suspension is sonicated with a ultrasonic disrupter (Branson) to disrupt the cells. The disrupted cell suspension is centrifuged to recover a soluble protein solution. The resultant protein solution is added to a reaction mixture containing at final concentrations of 100 mM Tris-HCl (pH 7.4), 5 mM DTT (dithiothreitol), 0.01% BSA, 200 μM sucrose, 5 mM galactinol and 31.7 μM [^{14}C] sucrose. The reaction mixture is

incubated at 37°C, followed by addition of 1.5 times in volume of ethanol and stirring. Insoluble materials are remove by centrifugation, the supernatant is spotted on, for example, a HPTLC cellulose thin layer chromatography plate (Merch, HPTLC plates cellulose) and then the plate is developed with n-butanol-pyridine-water-acetic acid (60 : 40 : 30 : 3). The developed plate is dried and analyzed with an imaging analyzer (FUJIX Bio-Image Analyzer BAS-2000II manufactured by Fuji Film) to determine [^{14}C] raffinose produced to measure the raffinose synthase activity.

In addition, the translated product as prepared above can also be used as an antigen for producing an antibody. The antibody thus produced can be used for, for example, detection and determination of the gene of the present invention in a crude protein extract prepared from an organism such as a plant.

When the host is a plant, the vector into which the gene of the present invention is inserted can be introduced into plant cells by a conventional means such as *Agrobacterium* infection method (JP-B 2-58917 and JP-A 60-70080), electroporation into protoplasts (JP-A 60-251887 and JP-B 5-68575) or particle gun method (JP-A 5-508316 and JP-A 63-258525). The plant cell transformed by the introduction of the vector can be selected on the basis of a selection marker, for example, resistance to an antibiotic such as kanamycin or hygromycin. From the plant cell thus transformed, a transformed plant can be regenerated by a conventional plant cell cultivation method,

for example, described in "Plant Gene Manipulation Manual (How to Produce Transgenic Plants)" written by Uchimiya, 1990, Kodan-sha Scientific (ISBN 4-06-153513-7), pp. 27-55. Furthermore, the collection of seeds from the transformed plant also makes it possible to prolify the transformed plant. In addition, crossing between the transformed plant obtained and the non-transformed plant makes it possible to produce progenic plants with the characters of the transformed plant.

As gene engineering techniques in soybean, basically, the above general techniques can be employed. More specifically, "transformation of soybean plant strain by particle gun" described in EP 301749, gene introduction methods. for example, described in Torisky, R.S., Kovacs, L., Avdiushko, S., Newman, J.D., Hunt, A.G. and Collins, G.B., "Development of a binary vector system for plant transformation based on the supervirulent Agrobacterium tumefaciens strain Chry5", Plant Cell Rep., (1997), 17, p. 102-108, etc. can be employed.

As gene engineering techniques in a Chenopdiaceae plant, basically, the above general techniques can be employed. More specifically, gene introduction methods, for example, described in M. Mannerlof, S. Tuvesson, P. Steen and P. Tenning, "Transgenic sugar beet tolerant to glyphosate", Euphytica (1997), 94, p 83-91, B.K. Konwar, "Agrobacterium tumefacience-Mediated Genetic Transformation of Sugar Beet (Beta vulgaris L.)", J. Plant Biochemistry & Biotechnology (1994), 3, p. 37-41 can be employed.

As gene engineering techniques in a Cruciferae plant, basically, the above general techniques can be employed. More specifically, the gene introduction can be carried out according to a method, for example, described in J. Fry, A. Barnason and R.B. Horsch, "Transformation of Brassica napus with Agrobacterium tumefaciens based vectors", Plant Cell Reports (1987), 6, 321-325.

For example, when gene introduction is carried out by *Agrobacterium* infection method, first, the above-described expression nucleic acid of the present invention is inserted into a binary vector. The resultant vector can be introduced into, for example, *Agrobacterium tumefaciens* LBA 4404 strain which has been converted into a competent state by treatment with calcium chloride. A transformant can be selected by an appropriate selection method according to the selection marker gene of the vector, for example, cultivation of a strain containing the vector in a culture medium containing an antibiotic in case that the selection marker gene is that giving resistance to the antibiotic such as kanamycin. The resultant transformed *Agrobacterium* strain can be culture in a liquid culture medium, for example, LB medium.

Soybean, a Chenopdiaceae plant or a Cruciferae plant can be transformed by using thus obtained *Agrobacterium* transformant culture broth as described below. For example, seeds from soybean, beet, rapeseed or mustard is sowed aseptically in, for example, 1/2 MS medium containing 2% sucrose and 0.7%

agar. After about 1 week, cotyledons and petioles of the germinated plant are cut off with a scalpel aseptically and transplanted in, for example, MS medium containing 3% sucrose, 0.7% agar, 4.5 μM BA, 0.05 μM 2,4-D and 3.3 μM $AgNO_3$ and cultured for one day. The cotyledons and petioles thus precultured are transferred to 1000-fold dilution of the above *Agrobacterium* culture broth and allowed to stand for 5 minutes. The cotyledons and petioles are transferred to the same medium as that of the preculture again and cultured for about 3 to 4 days. The cotyledones and petioles thus cultured are transferred to, for example, MS medium containing 3% sucrose, 4.5 μM BA, 0.05 μM 2,4-D, 3.3 μM $AgNO_3$ and 500 mg/liter cefotaxim, followed by shaking for 1 day to remove microbial cells. The resultant cotyledons and petioles are transferred to, for example, MS medium containing 3% sucrose, 0.7% agar, 4.5 μM BA, 0.05 μM 2,4-D, 3.3 μM $AgNO_3$ 100 mg/liter cefotaxim and 20 mg/liter kanamycin, followed by culturing for 3 to 4 weeks. Then, the cotyledons and petioles are transferred to, for example, MS medium containing 3% sucrose, 0.7% agar, 4.5 μM BA, 0.05 μM 2,4-D, 100 mg/liter cefotaxim and 20 mg/liter kanamycin and cultured. Culture in this medium is continued with subculturing every 3 to 4 weeks. When a shoot are regenerated, it is subcultured in, for example, MS medium containing 3% sucrose, 0.7% agar and 20 mg/liter kanamycin for 3 to 4 weeks. When the plant makes roots, it is transferred to vermiculite-peat moss (1 : 1) and acclimatized by culturing at 21 to 22°C under day and

night conditions of 12 hours: 12 hours = day time : night. As the plant grows, it is transferred to appropriate cultivation soil to culture the plant. A genomic DNA is extracted from the leaf of the regenerated plant according to the above method and PCR is carried out by using as primers having partial nucleotide sequences of the expression nucleic acid of the present invention to confirm the insertion of the gene of the present invention into the plant.

As described hereinabove, by introducing the gene of the present invention into a plant, for example, soybean, a Chenopdiaceae plant or a Cruciferae plant, it is possible to vary the expression level and activity of raffinose synthase in the plant to control the amount of raffinose family oligosaccharides in the plant. The gene of the present invention is useful in techniques for varying the expression level and activity of raffinose synthase in soybean, a Chenopdiaceae plant or a Cruciferae plant on the basis of gene homology, for example, techniques such as homologous recombination and antisense technique, cosuppression and the like.

The following examples further illustrate the present invention in detail but are not to be construed to limit the scope of the present invention.

Example 1

Preparation of cDNA Derived from Soybean

About 2 g of immature seeds of soybean (*Glycine max*) Williams82 were frozen in liquid nitrogen and then ground with

a mortar, to which 20 ml of Isogen (Nippon Gene) was added, and the mixture was further thoroughly ground. The ground material was transferred into a centrifugation tube, to which 4 ml of chloroform was added, and the mixture was stirred with a vortex mixer and then centrifuged at 6,500 x g for 10 minutes at 4°C. The water layer was collected, to which 10 ml of isopropanol was added, and the mixture was stirred and then centrifuged at 6,500 x g for 10 minutes at 4°C. The resulting precipitate was washed with 10 ml of 70% ethanol and then dissolved in 1 ml of elution buffer (10 mM Tris-HCl/pH 7.5, 1 mM EDTA, 0.1% SDS). The solution was allowed to stand at 60°C for 10 minutes and then centrifuged at 10,000 x g for 1 minute to remove insoluble matter. To the resulting supernatant was added an equivalent volume of Oligotex-dT30 (Takara), and the mixture was stirred and then allowed to stand at 65°C for 5 minutes. Further, the mixture was placed on ice and allowed to stand for 3 minutes, to which 200 μl of 5 M NaCl was added, and the mixture was mixed and then allowed to stand at 37°C for 10 minutes. The mixture was then centrifuged at 10,000 x g for 3 minutes at 4°C. The precipitate was collected and then suspended in 1 ml of TE buffer, and the suspension was allowed to stand at 65°C for 5 minutes. Further, the suspension was placed on ice and then allowing to stand for 3 minutes, followed by centrifugation at 10,000 x g for 3 minutes at 4°C to remove precipitate.

To the resulting supernatant were added 100 μl of 3M

sodium acetate and 2 ml of ethanol to precipitate and collect RNA. The collected RNA was washed twice with 70% ethanol and then dissolved in 20 μl of sterilized water, which was used for the subsequent cDNA synthesis. The amount of RNA obtained was determined by the measurement of absorbance at 260 nm.

For the cDNA synthesis, First Strand Synthesis Kit for RT-PCR (Amersham) and cDNA Synthesis Kit (Takara) were used, and all operations were made according to the protocol attached to kits.

Example 2

Cloning of Raffinose Synthase Gene from Soybean cDNA

PCR was carried out by using the cDNA obtained from immature seeds of soybean (*Glycine max*) Williams82 in Example 1 as a template and the primers designed on the basis of the amino acid sequence of SEQ ID No: 1, i.e., primers having nucleotide sequences shown in List 4 (SEQ ID NOS: 21 and 22) below to amplify a DNA fragment. The PCR was carried out with Gene Amp PCR Systems 2400 and DNA Thermal Cycler Model 480 of Perkin-Elmer using Advantage KlenTaq cDNA Kit of Clontech. The reaction was carried out by repeating the cycle for maintaining at 94°C for 1 minute, at 50°C for 3 minutes and then at 72°C for 3 minutes 40 times to amplify the DNA fragment. The amplified DNA fragment was cloned with TA cloning kit (Invitrogen), followed by sequence reaction using ABI PRISM Dye Terminator Cycle Sequencing Ready Reaction Kit of Perkin-Elmer and analysis of the nucleotide sequence with a 373S DNA sequencer of ABI. Based on this sequence, primers

having nucleotide sequences shown in List 5 below were synthesized. The synthesis of cDNA was carried out with Marathon Kit of Clontech using mRNA obtained from leaves of soybean Williams82 in Example 1. The cDNA obtained was ligated to an adapter contained in the kit with ligase. These operations were carried out according to the protocol attached to the kit. By using the adapter-ligated cDNA thus prepared, PCR was carried out with the primers shown in List 5 (SEQ ID NO: 23) according to the same manner as the above. The nucleotide sequence in terminal region of the gene was analyzed according to the protocol attached to the Marathon Kit of Clontech. As a result, the nucleotide sequence of SEQ ID NO: 2 was determined.

List 4

4-5-F primer (SEQ ID NO: 21):

cgatggatgg giaaittiat icaiccigai tgggaiatgt t 41mer

4-6-RV primer (SEQ ID NO: 22):

ggccacatit tiacia(ag)icc iatiggigci aa 32mer

List 5

5-SC-2 (SEQ ID NO: 23):

tgttactagg cgaaacaaga gtagctctga 30mer

Example 3

Preparation of cDNA derived from Chenopdiaceae Plant

About 2 g of leaves of beet (*Beta vulgaris*: haming) was frozen in liquid nitrogen and then ground with a mortar, to which 20 ml of Isogen (Nippon Gene) was added, and the mixture was further thoroughly ground. The ground material was

transferred into a centrifugation tube, to which 4 ml of chloroform was added, and the mixture was stirred with a vortex mixer and then centrifuged at 6,500 x g for 10 minutes at 4°C. The water layer was collected, to which 10 ml of isopropanol was added, and the mixture was stirred and then centrifuged at 6,500 x g for 10 minutes at 4°C. The resulting precipitate was washed with 10 ml of 70% ethanol and then dissolved in 180 μl of DEPC-treated sterilized water. The solution was allowed to stand at 55°C for 5 minutes and 20 μl of 5 M NaCl was added thereto. The resulting solution was purified using BIOMAG mRNA PURIFICATION KIT (PerSeptive Biosystems: Catalog No. 8-MB4003K).

To the resulting mRNA solution were added 3M sodium acetate and ethanol, and RNA was precipitated and collected. The collected RNA was washed twice with 70% ethanol and then dissolved in 20 μl of sterilized water, which was used for the subsequent cDNA synthesis. The amount of RNA obtained was determined by the measurement of absorbance at 260 nm.

For the cDNA synthesis, SMART PCR cDNA Synthesis Kit (Clontech) was used, and all operations were made according to the protocol attached to the kit.

Example 4

Analysis of Nucleotide Sequence of Raffinose Synthase Gene from Chenopdiaceae Plant

Synthetic DNA primers having the nucleotide sequences shown in List 6 (SEQ ID NOS: 24-27) below were synthesized. The PCR method was carried out with Gene Amp PCR Systems 2400

and DNA Thermal Cycler Model 480 of Perkin-Elmer using Advantage KlenTaq cDNA Kit of Clontech. The PCR was carried out with the above primers (SEQ ID NOS: 24-27) and cDNA of beet obtained in the above Example 3 by repeating the cycle for maintaining at 94°C for 1 minute, at 50°C for 3 minutes and then at 72°C for 3 minutes 40 times. As a result, the combinations of primers 6-3-F (SEQ ID NO: 24) and 6-8-RV (SEQ ID NO: 25) and primers 6-10-F (SEQ ID NO: 26) and 6-6-RV (SEQ ID NO: 27) gave an amplification of about 0.3 kb and 0.6 kb bands, respectively. The amplified DNA fragments were cloned with TA cloning kit (Invitrogen), followed by sequence reaction using ABI PRISM Dye Terminator Cycle Sequencing Ready Reaction Kit of Perkin-Elmer and analysis of nucleotide sequence with a 373S DNA sequencer of ABI. Based on the resulting nucleotide sequences, synthetic DNA primers having nucleotide sequences shown in List 7 (SEQ ID NOS: 28-35) below were prepared and PCR was carried out using cDNA obtained from beet in Example 3 in the same manner as above. As a result, DNA having the nucleotide sequence of SEQ ID NO: 4 was finally obtained from cDNA of beet.

List 6

6-3-F (SEQ ID NO: 24):

cgaggiggit gicciccigg ittigtiati atigaigaig gitggca 47mer

6-8-RV (SEQ ID NO: 25):

at(t/c)tt(a/g)tcia cigcia(a/g)(a/g)tc (t/c)tccatigt 29mer

6-10-F (SEQ ID NO: 26):

ggiacitait gg(c/t)ticaigg itgicaiatg gticaitg 38mer

6-6-RV (SEQ ID NO: 27):

ggccacatit tiacia(a/g)icc iatiggigci aa 32mer

List 7

7-Sb-1 (SEQ ID NO: 28):

atctatttgt catgacgatg atccga 26mer

7-Sb-2RV (SEQ ID NO: 29):

ggccctcatt cccatattgg gatgatcctc 30mer

7-Sb-3RV (SEQ ID NO: 30):

aagcatgcca aacatacaca tgctcaacag 30mer

7-Sb-4RV (SEQ ID NO: 31):

agacccgggg aaagctttgg ggttactact 30mer

7-Sb-5 (SEQ ID NO: 32):

tggatgggaa actttataca ccctgact 28mer

7-Sb-6 (SEQ ID NO: 33):

gacatgttcc catctacaca cccttgtg 28mer

7-Sb-7 (SEQ ID NO: 34):

ccaatttatg ttagtgatgt tgttggcaag 30mer

7-Sb-8RV (SEQ ID NO: 35):

tcgactccca gggtagaatt gtcatc 26mer

Example 5

Preparation of cDNA Derived from Cruciferae Plant

About 2 g of leaves of mustard (*Brassica juncea*) was frozen in liquid nitrogen and then ground with a mortar, to which 20 ml of Isogen (Nippon Gene) was added, and the mixture was further thoroughly ground. The ground material was transferred into a centrifugation tube, to which 4 ml of chloroform

was added, and the mixture was stirred with a vortex mixer and then centrifuged at 6,500 x g for 10 minutes at 4°C. The water layer was collected, to which 10 ml of isopropanol was added, and the mixture was stirred and then centrifuged at 6,500 x g for 10 minutes at 4°C. The resulting precipitate was washed with 10 ml of 70% ethanol and then dissolved in 180 μl of DEPC-treated sterilized water. The solution was allowed to stand at 55°C for 5 minutes and to which 20 μl of 5 M NaCl was added. The resulting solution was purified using BIOMAG mRNA PURIFICATION KIT (PerSeptive Biosystems: Catalog No. 8-MB4003K).

To the resulting mRNA solution were added 3M sodium acetate and ethanol, and RNA was precipitated and collected. The collected RNA was washed twice with 70% ethanol and then dissolved in 20 μl of sterilized water, which was used for the subsequent cDNA synthesis. The amount of RNA obtained was determined by the measurement of absorbance at 260 nm.

For the cDNA synthesis, SMART PCR cDNA Synthesis Kit (Clontech) was used, and all operations were carried out according to the protocol attached to the kit.

In the same manner as described in the above, mRNA was purified from immature seeds of rapeseed Westar (*Brassica napus*) and cDNA was synthesized.

Example 6

Isolation and Nucleotide Sequence Analysis of Raffinose Synthase Gene derived from Cruciferae Plant

DNA primers having the nucleotide sequences shown in

List 8 (SEQ ID NOS: 36 and 37) below were synthesized. PCR was carried out with Gene Amp PCR Systems 2400 and DNA Thermal Cycler Model 480 of Perkin-Elmer using Advantage KlenTaq cDNA Kit of Clontech. The PCR was carried out with the above primers and cDNA of mustard obtained in Example 5 by repeating the cycle for maintaining at 94°C for 1 minute, at 50°C for 3 minutes and then at 72°C for 3 minutes 40 times. The reaction products were analyzed by agarose gel electrophoresis. As a result, an amplification of about the 1.2 kb bands was detected. The amplified DNA fragment was cloned with TA cloning kit (Invitrogen), followed by sequence reaction using ABI PRISM Dye Terminator Cycle Sequencing Ready Reaction Kit of Perkin-Elmer and nucleotide sequence analysis with a 373S DNA sequencer of ABI. Based on the resulting nucleotide sequence, synthetic primers having the nucleotide sequences shown in List 9 (SEQ ID NOS: 38 and 39) below were prepared and PCR was carried out using cDNAs from mustard (*Brassica juncea*) and rapeseed Westar (*Brassica napus*) obtained in Example 5 according to the same manner as the above. As a result, the nucleotide sequence represented by the 749th to 1215th nucleotides of SEQ ID NO: 6 and by the 1st to 467th nucleotides of SEQ ID NO: 8 were finally determined for cDNA from mustard (*Brassica juncea*) and rapeseed Westar (*Brassica napus*), respectively.

List 8

8-#1 (SEQ ID NO: 36):

cgattiaaig titggtggac iacicaitgg gtigg 35mer

8-#10RV (SEQ ID NO: 37):

caitgiacca titgicaicc itgia(ag)ccai taigticc 38mer

List 9

9-primer-1 (SEQ ID NO: 38):

gttagggttc atatgaacac cttcaagctc 30mer

9-primer-2RV (SEQ ID NO: 39):

caacggcgag atcttgcatc gtcaac 26mer

Example 7

Nucleotide Sequence Analysis of Raffinose Synthase Full-Length Gene Derived from Cruciferae Plant

Based on the nucleotide sequences obtained in Example 6, DNA primers having the nucleotide sequences shown in List 10 (SEQ ID NOS: 40-46) below were synthesized. The cDNAs from mustard (*Brassica juncea*) and rapeseed Westar (*Brassica napus*) obtained in the same manner as described in Example 5 were ligated to adapters contained in Marathon Kit of Clontech. By using the adapter-ligated cDNAs thus prepared, PCR was carried out with primers shown in List 10. 10-B-2RV (SEQ ID NO: 40), 10-B-3RV (SEQ ID NO: 41) and 10-B-4RV (SEQ ID NO: 42) primers were used for nucleotide analysis of 5'-termini, and 10-B-1 (SEQ ID NO: 43), 10-B-8 (SEQ ID NO: 44), 10-B-7 (SEQ ID NO: 45) and 10-B-6 (SEQ ID NO: 46) primers were used for nucleotide analysis of 3'-termini. The nucleotide sequences were analyzed according to the protocol attached to the Marathon Kit of clontech. As a result, the nucleotide sequence of SEQ ID NO: 6 and SEQ ID NO: 8 were determined from mustard (*Brassica juncea*) and rapeseed

Westar (*Brassica napus*), respectively.

List 10

10-B-2RV (SEQ ID NO: 40):

ggattcgaca caaaccgcca cgtcatcgtc 30mer

10-B-3RV (SEQ ID NO: 41):

ccacgtgcac cacccgaact tatcgac 27mer

10-B-4RV (SEQ ID NO: 42):

aacatcgata ccatcggagt catgtccaat 30mer

10-B-1 (SEQ ID NO: 43):

gttagggttc atatgaacac cttcaagctc 30mer

10-B-8 (SEQ ID NO: 44):

tctacgtctg gcacgcgctt tgcggctac 29mer

10-B-7 (SEQ ID NO: 45):

gttgacgtca tccacatatt ggagatgttg t 31mer

10-B-6 (SEQ ID NO: 46):

gttatcgcta gcatggagca ctgtaatga 29mer

Example 8

Construction of Expression Vectors in Plant for Raffinose Synthase Gene Derived from Cruciferae Plant

Based on the nucleotide sequence of raffinose synthase gene from mustard obtained in Example 7, DNA primers having the nucleotide sequences shown in List 11 (SEQ ID NOS: 47 and 48) were prepared. By using cDNA of mustard, PCR was carried out in the same manner as described in Example 6. The amplified DNA fragment was digested with SacI. The DNA fragment thus digested was ligated to the vector pBI121(-) previously digested

with SacI by using Ligation Kit (Takara). Plasmid pBI121 (Clontech) were digested with BamHI and SacI, and ligated to linkers shown in List 12 (SEQ ID NOS: 49 and 50) to prepare the vector pBI121(-). The vector thus obtained was analyzed by a restriction map and PCR using primers having nucleotide sequences shown in List 13 (SEQ ID NOS: 51-53), and confirmed the direction of inserted raffinose synthase gene. The vector whose raffinose synthase gene from mustard was inserted in the expressible direction was designated BjRS-Sac(+)-121 and the one whose raffinose synthase gene from mustard was inserted in the reverse direction was designated BjRS-Sac(-)-121.

List 11

11-SacI-BjN (SEQ ID NO: 47):

aacgagctca atccaaaatc tcatcaaata atcgc 35mer

11-SacI-BjintRV (SEQ ID NO: 48):

acaatagttg agggcggaag agtag 25mer

List 12

12-BamSac-(+)linker (SEQ ID NO: 49):

gatcgagctc gtgtcggatc cagct 25mer

12-BamSac-(-)linker (SEQ ID NO: 50):

ggatccgaca cgagctc 17mer

List 13

13-35S-3 (SEQ ID NO: 51):

cctcctcgga ttccattgcc cagctatctg 30mer

13-B-2RV (SEQ ID NO: 52):

ggattcgaca caaaccgcca cgtcatcgtc 30mer

13-B-8 (SEQ ID NO: 53):

tctacgtctg gcacgcgctt tgcggctac 29mer

Example 9

Transformation with Raffinose Synthase Gene Derived from Cruciferae Plant

The vectors BjRS-Sac(+)-121 and BjRS-Sac(-)-121 prepared in Example 8 were used for the transformation of mustard (*Brassica juncia*) by the *Agrobacterium* infection method.

Agrobacterium tumefacients (strain LBA4404 having rifampicin and streptomycin resistance) previously converted into a competent state by calcium chloride treatment was transformed independently with two plasmids BjRS-Sac(+)-121 and BjRS-Sac(-)-121 prepared in Example 8. The transformants were selected on LB medium containing 50 μg/ml rifampicin and 25 μg/ml kanamycin by utilizing the kanamycin resistant character conferred by the kanamycin resistant gene (neomycin phosphotransferase, NPTII) of the introduced plasmids.

The transformant Agrobacterium obtained (*Agrobacterium tumefaciens* strain LBA4404: rifampicin and streptomycin resistant) was cultured on LB medium containing 50 μg/ml rifampicin and 25 μg/ml kanamycin at 28°C for a whole day and night, and the culture was used for the transformation of mustard by the method described below.

The seeds of mustard were aseptically sowed on 1/2 MS medium containing 2% sucrose and 0.7% agar. After one week, cotyledons and petioles of sprouting plants were cut out with

a scalpel, and transferred to MS medium containing 3% sucrose, 0.7% agar, 4.5 μM BA, 0.05 μM 2.4-D and 3.3 μM $AgNO_3$, followed by preculture for 1 day. The precultured cotyledons and petioles were transferred in a 1000-fold dilution of the Agrobacterium culture broth and allowed to stand for 5 minutes. The cotyledons and petioles were transferred again to the same medium as used in the preculture, and cultured for 3 to 4 days. The cultured cotyledons and petioles were transferred to MS medium containing 3% sucrose, 4.5 μM BA, 0.05 μM 2.4-D, 3.3 μM $AgNO_3$ and 500 mg/l cefotaxim, and shaken for 1 day to remove microbial cells. The cotyledons and petioles thus treated were transferred to MS medium containing 3% sucrose, 0.7% agar, 4.5 μM BA, 0.05 μM 2.4-D, 3.3 μM $AgNO_3$, 100 mg/l cefotaxim and 20 mg/l kanamycin, and cultured for 3 to 4 weeks. The cotyledons and petioles were transferred to MS medium containing 3% sucrose, 0.7% agar, 4.5 μM BA, 0.05 μM 2.4-D, 100 mg/l cefotaxim and 20 mg/l kanamycin, and cultivated. The cultivation on this medium was continued with subculturing at intervals of 3 to 4 weeks. When shoots are began to regenerate, these shoots are subcultured on MS medium containing 3% sucrose, 0.7% agar and 20 mg/l kanamycin, and cultivated for 3 to 4 weeks. The rooting plants are transferred to vermiculite : peat moss = 1 : 1, and cultivated at 21°C to 22°C in a cycle of day/night = 12 hours : 12 hours. With the progress of plant body growth, the plants are grown with cultivation soil.

BRIEF DESCRIPTION OF SEQUENCES

SEQ ID NO: 1 shows an amino acid sequence of a raffinose synthase protein encoded by the raffinose synthase gene of the present invention.

SEQ ID NO: 2 shows a nucleotide sequence of the raffinose synthase gene of the present invention.

SEQ ID NO: 3 shows an amino acid sequence of a raffinose synthase protein encoded by the raffinose synthase gene obtained from beet.

SEQ ID NO: 4 shows a cDNA nucleotide sequence of the raffinose synthase gene obtained from beet.

SEQ ID NO: 5 shows an amino acid sequence of a raffinose synthase protein encoded by the raffinose synthase gene obtained from mustard.

SEQ ID NO: 6 shows a cDNA nucleotide sequence of the raffinose synthase gene obtained from mustard.

SEQ ID NO: 7 shows an amino acid sequence of a raffinose synthase protein encoded by the raffinose synthase gene obtained from rapeseed.

SEQ ID NO: 8 shows a cDNA nucleotide sequence of the raffinose synthase gene obtained from rapeseed.

List 1:

The nucleotide sequences shown in List 1 (SEQ ID NOS: 9 and 10) are examples of the typical primers used in the amplification of a DNA fragment having a raffinose synthase gene. All of these sequences are based on the nucleotide sequence

of SEQ ID NO: 2. Primer 11 (SEQ ID NO: 9) is a sense primer and Primer 12 (SEQ ID NO: 10) is an antisense primer. Depending upon the purpose, recognition sequences for suitable restriction enzymes can be added to the 5'-termini of these nucleotide sequences.

List 2:

The nucleotide sequences shown in List 2 (SEQ ID NOS: 11-14) are examples of the typical primers used in the amplification of a cDNA of a raffinose synthase gene. Primer 21 (SEQ ID NO: 11) is a sense primer corresponding to the 5'-terminus of the beet-derived raffinose synthase gene. Primer 22 (SEQ ID NO: 12) is an antisense primer corresponding to the 3'-terminus. Depending upon the purpose, recognition sequences for suitable restriction enzymes can be added to the 5'-termini of these nucleotide sequences.

Primer 23 (SEQ ID NO: 13) is a sense primer corresponding to the N-terminus of the open reading frame, and primer 24 (SEQ ID NO: 14) is an antisense primer corresponding to the C-terminus.

List 3:

Among the nucleotide sequences shown in List 3 (SEQ ID NOS: 15-20), primers 31 (SEQ ID NO: 15) and 32 (SEQ ID NO: 16) are typical primers used in the amplification of a DNA having the partial nucleotide sequence of a raffinose synthase gene. Primer 31 (SEQ ID NO: 15) is a sense primer used in the amplification of a DNA having the partial nucleotide sequence of a raffinose synthase gene from mustard and rapeseed and primer 32 (SEQ ID

NO: 16) is an antisense primer. Depending upon the purpose, recognition sequences for suitable restriction enzymes can be added to the 5'-termini of these nucleotide sequences.

Primers 33 (SEQ ID NO: 17) and 34 (SEQ ID NO: 18) are the typical primers used in the amplification of a cDNA of a raffinose synthase gene of mustard. Primers 33 (SEQ ID NO: 17) and 34 (SEQ ID NO: 18) are both based on the nucleotide sequence of raffinose synthase gene in the non-translated region. Primer 33 (SEQ ID NO: 17) is a sense primer corresponding to the 5'-terminal non-translated region of the mustard-derived raffinose synthase gene. Primer 34 (SEQ ID NO: 18) is an antisense primer corresponding to the 3'-terminal non-translated region.

Primers 35 (SEQ ID NO: 19) and 36 (SEQ ID NO: 20) are typical primers used in the amplification of an open reading frame coding for the amino acid sequence of a raffinose synthase protein in the cDNA of a raffinose synthase gene. Primer 35 (SEQ ID NO: 19) is a sense primer corresponding to the 5'-terminus of the above open reading frame. Primer 36 (SEQ ID NO: 20) is an antisense primer corresponding to the 3'-terminus. Depending upon the purpose, recognition sequences for suitable restriction enzymes can be added to the 5'-termini of these nucleotide sequences.

List 4:

The nucleotide sequences shown in List 4 (SEQ ID NOS: 21 and 22) are of the primers used in the cloning of a DNA fragment having the present raffinose synthase gene. The base represented

by the symbol "i" is inosine. The bases shown in parentheses mean that a mixture of those bases is used in the synthesis. The symbol "RV" as used after the primer number means that the primer referred to by this symbol has an antisense sequence.

List 5:

The nucleotide sequence shown in List 5 (SEQ ID NO: 23) is of the primer used in the analysis of a nucleotide sequence of the present raffinose synthase gene. 5-SC-2 (SEQ ID NO: 23) is used in the analysis of the present nucleotide sequence in the 3'-terminal region.

List 6

The nucleotide sequences shown in List 6 (SEQ ID NOS: 24-27) are of the primers used in the analysis of the present raffinose synthase gene of beet. The base represented by the symbol "i" is inosine. The bases shown in parentheses mean that a mixture of those bases was used in the synthesis. The symbol "RV" as used after the primer number means that the primer referred to by this symbol has an antisense sequence.

List 7

The nucleotide sequences shown in List 7 (SEQ ID NOS: 28-35) are of the primers synthesized on the partial nucleotide sequences of the beet raffinose synthase gene. The symbol "RV" as used after the primer number means that the primer referred to by this symbol has an antisense sequence.

List 8:

The nucleotide sequences shown in List 8 (SEQ ID NOS:

36 and 37) are of the primers used in the analysis of the cDNA nucleotide sequence of a raffinose synthase gene of mustard. The base represented by the symbol "i" is inosine. The bases shown in parentheses mean that a mixture of those bases. The symbol "RV" as used after the primer number means that the primer referred to by this symbol has an antisense sequence.

List 9:

The nucleotide sequences shown in List 9 (SEQ ID NOS: 38 and 39) are of the primers synthesized on the partial nucleotide sequences of the mustard raffinose synthase gene. The symbol "RV" as used after the primer number means that the primer referred to by this symbol has an antisense sequence.

List 10:

The nucleotide sequences shown in List 10 (SEQ ID NOS: 40-46) are of the primers used in the analysis of the nucleotide sequences of raffinose synthase gene of mustard and rapeseed. The symbol "RV" as used after the primer number means that the primer referred to by this symbol has an antisense sequence.

List 11:

The nucleotide sequences shown in List 11 (SEQ ID NOS: 47 and 48) are of the primers used in the amplification of 5'-terminal region of a mustard raffinose synthase gene. 11-SacI-BjN (SEQ ID NO: 47) is a primer whose SacI restriction site is added to the nucleotide sequence represented by the 4th to 29th nucleotides in SEQ ID NO: 6. 11-SacI-BjintRV (SEQ ID NO: 48) is an antisense primer having a nucleotide sequence

corresponding to the nucleotide sequence represented by the 1164th to 1188th nucleotides in SEQ ID NO: 6.

List 12:

The nucleotide sequences shown in List 12 (SEQ ID NOS: 49 and 50) are of the adapters added to a mustard cDNA. These synthetic DNA take a double-stranded form when mixed together because they are complementary strands. This adapter has cohesive ends of cleavage sites for the restriction enzymes BamHI and SacI on both termini, and contains the restriction sites for the restriction enzymes BamHI and SacI in the double-stranded region.

List 13:

The nucleotide sequences shown in List 13 (SEQ ID NOS: 51-53) are of the primers used in the confirmation of inserting direction of the mustard-derived raffinose synthase gene. 13-35S-3 (SEQ ID NO: 51) is a primer of sense to 35S promoter. 13-B-2RV (SEQ ID NO: 52) is an antisense primer having the nucleotide sequence represented by the 593rd to 622nd nucleotides of SEQ ID NO: 6, 13-B-8 (SEQ ID NO: 53) is a sense primer having the nucleotide sequence represented by the 1110th to 1138th nucleotides in SEQ ID NO: 6.

As described hereinabove, according to the present invention, it is possible to provide raffinose synthase genes which can be utilized in techniques for varying expression level and activity of raffinose synthase in plants.

SEQUENCE LISTING FREE TEXT

SEQ ID NO: 9 to SEQ ID NO: 20: Designed oligonucleotide primer to obtain raffinose synthase gene.

SEQ ID NO: 21 and SEQ ID NO: 22: Designed oligonucleotide primer to obtain raffinose synthase gene, n is i, r is a or g.

SEQ ID NO: 23: Designated oligonucleotide primer to obtain raffinose synthase gene.

SEQ ID NO: 24 to SEQ ID NO: 27: Designed oligonucleotide primer to obtain raffinose synthase gene, n is i, y is t or c, r is a or g.

SEQ ID NO: 28 to SEQ ID NO: 35: Designed oligonucleotide primer to obtain raffinose synthase gene.

SEQ ID NO: 36 and SEQ ID NO: 37: Designed oligonucleotide primer to obtain raffinose synthase gene, n is i, r is a or g.

SEQ ID NO: 38 to SEQ ID NO: 48: Designed oligonucleotide primer to obtain raffinose synthase gene.

SEQ ID NO: 49 and SEQ ID NO: 50: Designed oligonucleotide linker to obtain raffinose synthase gene.

SEQ ID NO: 51 to SEQ ID NO: 53: Designed oligonucleotide primer to confirm direction of the inserted raffinose synthase gene.

SEQUENCE LISTING

<110> Eijiro WATANABE et al.; Sumitomo Chemical Company Limited

<120> Raffinose Synthase Genes and Their Use

<150> JP 10/120550
<151> 1998-04-30

<150> JP 10/120551
<151> 1998-04-30

<150> JP 10/345590
<151> 1998-12-04

<150> JP 10/351246
<151> 1998-12-10

<160> 53

<210> 1
<211> 265
<212> PRT
<213> Glycine max

<400> 1

```
Gln Ser Asp His Ala Cys Ala Glu Phe His Ala Ala Ser Arg Ala Ile
 1               5                  10                  15
Ser Gly Gly Pro Ile Tyr Val Ser Asp Ser Val Gly Lys His Asn Phe
            20                  25                  30
Lys Leu Leu Lys Lys Leu Val Leu Pro Asp Gly Ser Ile Leu Arg Cys
        35                  40                  45
Gln His Tyr Ala Leu Pro Thr Arg Asp Cys Leu Phe Val Asp Pro Leu
    50                  55                  60
His Asp Gly Lys Thr Met Leu Lys Ile Trp Asn Leu Asn Lys Cys Ser
65                  70                  75                  80
Gly Val Leu Gly Leu Phe Asn Cys Gln Gly Gly Gly Trp Cys Pro Val
                85                  90                  95
```

Thr Arg Arg Asn Lys Ser Ser Ser Asp Tyr Ser His Ser Val Thr Cys
100 105 110
Phe Ala Ser Pro Gln Asp Ile Glu Trp Gly Lys Gly Lys His Pro Val
115 120 125
Cys Ile Lys Gly Val Asp Val Phe Ala Val Tyr Met Phe Lys Asp Asp
130 135 140
Lys Leu Lys Leu Leu Lys Tyr Thr Glu Ser Val Glu Val Ser Leu Glu
145 150 155 160
Pro Phe Ser Cys Glu Leu Leu Thr Val Ser Pro Val Val Ile Leu Pro
165 170 175
Arg Lys Ser Ile Gln Phe Ala Pro Ile Gly Leu Val Asn Met Leu Asn
180 185 190
Ser Gly Gly Ser Ile Met Ser Leu Glu Phe Asp Gln Gln Glu Asn Leu
195 200 205
Ala Arg Ile Gly Val Arg Gly His Gly Glu Met Arg Val Phe Ala Ser
210 215 220
Glu Lys Pro Glu Ser Val Lys Ile Asp Gly Glu Ser Val Glu Phe Asp
225 230 235 240
Tyr Val Asp Arg Thr Val Arg Leu Gln Val Ser Trp Pro Cys Ser Ser
245 250 255
Arg Leu Ser Val Val Glu Tyr Leu Phe
260 265

<210> 2
<211> 928
<212> DNA
<213> Glycine max

<220>
<221> CDS
<222> (2)...(799)

<400> 2
c caa tct gat cat gct tgt gcc gaa ttc cac gct gct tct aga gcc 46
Gln Ser Asp His Ala Cys Ala Glu Phe His Ala Ala Ser Arg Ala
5 10 15
att tct ggt gga cca att tat gta agc gac tct gtt gga aaa cac aac 94
Ile Ser Gly Gly Pro Ile Tyr Val Ser Asp Ser Val Gly Lys His Asn
20 25 30
ttc aag ttg ctt aag aag ctt gtt cta cct gat ggc tcc att ttg cgg 142

```
Phe Lys Leu Leu Lys Lys Leu Val Leu Pro Asp Gly Ser Ile Leu Arg
                35                  40                  45
tgt caa cat tat gca ctt ccc acc cga gac tgc tta ttt gta gat cct   190
Cys Gln His Tyr Ala Leu Pro Thr Arg Asp Cys Leu Phe Val Asp Pro
        50                  55                  60
tta cat gat ggg aaa aca atg ctc aaa att tgg aac ctc aat aaa tgt   238
Leu His Asp Gly Lys Thr Met Leu Lys Ile Trp Asn Leu Asn Lys Cys
    65                  70                  75
tcc ggg gtt ttg ggt ctg ttc aat tgc caa gga gga ggt tgg tgc cct   286
Ser Gly Val Leu Gly Leu Phe Asn Cys Gln Gly Gly Gly Trp Cys Pro
80                  85                  90                  95
gtt act agg cga aac aag agt agc tct gac tat tca cac tcc gtg act   334
Val Thr Arg Arg Asn Lys Ser Ser Ser Asp Tyr Ser His Ser Val Thr
                100                 105                 110
tgc ttt gca agt cct caa gac att gaa tgg ggc aaa ggg aag cac cca   382
Cys Phe Ala Ser Pro Gln Asp Ile Glu Trp Gly Lys Gly Lys His Pro
            115                 120                 125
gtt tgc atc aaa ggg gtg gac gta ttt gct gtg tac atg ttt aag gac   430
Val Cys Ile Lys Gly Val Asp Val Phe Ala Val Tyr Met Phe Lys Asp
        130                 135                 140
gac aag ttg aag ctg ctg aag tac aca gag agt gta gaa gtt tct ctt   478
Asp Lys Leu Lys Leu Leu Lys Tyr Thr Glu Ser Val Glu Val Ser Leu
    145                 150                 155
gag cct ttt agt tgt gag ctt ttg acc gtt tct cca gtg gtg atc tta   526
Glu Pro Phe Ser Cys Glu Leu Leu Thr Val Ser Pro Val Val Ile Leu
160                 165                 170                 175
ccc aga aaa tca atc caa ttt gcc cca att gga ttg gta aac atg ctc   574
Pro Arg Lys Ser Ile Gln Phe Ala Pro Ile Gly Leu Val Asn Met Leu
                180                 185                 190
aac tct ggg ggc tct att atg tca ttg gaa ttt gat caa cag gaa aat   622
Asn Ser Gly Gly Ser Ile Met Ser Leu Glu Phe Asp Gln Gln Glu Asn
            195                 200                 205
ttg gcg agg att ggg gtg aga gga cat ggg gaa atg agg gta ttt gca   670
Leu Ala Arg Ile Gly Val Arg Gly His Gly Glu Met Arg Val Phe Ala
        210                 215                 220
tca gag aag cca gag agt gtc aag att gat gga gaa tct gtg gaa ttt   718
Ser Glu Lys Pro Glu Ser Val Lys Ile Asp Gly Glu Ser Val Glu Phe
    225                 230                 235
gat tat gtt gat aga acc gtg agg ctc caa gtc tcg tgg cct tgt tct   766
Asp Tyr Val Asp Arg Thr Val Arg Leu Gln Val Ser Trp Pro Cys Ser
```

240 245 250 255
tcg agg ttg tcc gta gtc gag tat ttg ttc tga atcatgattt ggtgtccgag 819
Ser Arg Leu Ser Val Val Glu Tyr Leu Phe
260 265
agagccgtgt aatgttcaca taaactgact taagtgcatt aagcaaatcc accttaaata 879
atagtgcata actttgttcc aaaaaaaaaa aaaaaaaaaa aaaaaaaaa 928

<210> 3
<211> 783
<212> PRT
<213> Beta vulgaris L.

<400> 3
Met Ala Pro Ser Phe Ser Lys Glu Asn Ser Lys Thr Cys Asp Glu Val
5 10 15
Ala Asn His Asp Asp Cys Asn Thr Cys Pro Ile Ile Ser Leu Glu Glu
20 25 30
Ser Asn Phe Met Val Asn Gly His Val Ile Leu Ser Gln Val Pro Ser
35 40 45
Asn Ile Thr Ala Ile Ser Lys Met Gly Phe Asp Gly Leu Phe Val Gly
50 55 60
Phe Asp Ala Pro Glu Pro Lys Ala Arg His Val Val Ser Val Gly Gln
65 70 75 80
Leu Lys Gly Ile Pro Phe Met Ser Ile Phe Arg Phe Lys Val Trp Trp
85 90 95
Thr Thr His Trp Thr Gly Ser Asn Gly Arg Asp Leu Glu His Glu Thr
100 105 110
Gln Ile Leu Ile Leu Asp Lys Ser Asp Glu Gly Leu Gly Arg Pro Tyr
115 120 125
Ile Val Ile Leu Pro Leu Ile Glu Gly Pro Phe Arg Ala Ser Leu Gln
130 135 140
Pro Gly Ser Val Asp Asp Tyr Val Asp Ile Cys Val Glu Ser Gly Ser
145 150 155 160
Thr Lys Val Val Gly Asp Ser Phe Arg Ala Val Leu Tyr Ile Arg Ala
165 170 175
Gly Pro Asp Pro Phe Lys Leu Ile Lys Asp Thr Met Lys Glu Val Gln
180 185 190
Ala His Leu Gly Thr Phe Lys Leu Leu Asp Asp Lys Thr Pro Pro Gly
195 200 205
Ile Val Asp Lys Phe Gly Trp Cys Thr Trp Asp Ala Phe Tyr Leu Lys

```
    210                 215                 220
Val Glu Pro Tyr Gly Val Trp Glu Gly Val Lys Gly Leu Val Glu Asn
225                 230                 235                 240
Gly Val Pro Pro Gly Leu Val Leu Ile Asp Asp Gly Trp Gln Ser Ile
                245                 250                 255
Cys His Asp Asp Asp Pro Ile Thr Asp Gln Glu Gly Ile Asn Arg Thr
            260                 265                 270
Ser Ala Gly Glu Gln Met Pro Cys Arg Leu Ile Lys Tyr Glu Glu Asn
        275                 280                 285
Phe Lys Phe Arg Asp Tyr Lys Ser Pro Asn Ile Met Gly His Glu Asp
    290                 295                 300
His Pro Asn Met Gly Met Arg Ala Phe Val Arg Asp Leu Lys Glu Glu
305                 310                 315                 320
Phe Lys Thr Val Glu His Val Tyr Val Trp His Ala Phe Thr Gly Tyr
                325                 330                 335
Trp Gly Gly Val Arg Pro Asn Val Pro Gly Leu Pro Glu Ala Gln Val
            340                 345                 350
Val Thr Pro Lys Leu Ser Pro Gly Leu Glu Met Thr Met Glu Asp Leu
        355                 360                 365
Ala Val Asp Lys Ile Val Asn Asn Gly Ile Gly Leu Val Gln Pro Asp
    370                 375                 380
Lys Ala Gln Glu Leu Tyr Glu Gly Leu His Ser His Leu Glu Asn Cys
385                 390                 395                 400
Gly Ile Asp Gly Val Lys Val Asp Val Ile His Leu Leu Glu Met Met
                405                 410                 415
Ala Glu Asp Tyr Gly Gly Arg Val Glu Leu Ala Lys Thr Tyr Tyr Lys
            420                 425                 430
Ala Ile Thr Glu Ser Val Arg Lys His Phe Lys Gly Asn Gly Val Ile
        435                 440                 445
Ala Ser Met Glu Gln Cys Asn Asp Phe Met Leu Leu Gly Thr Glu Thr
    450                 455                 460
Ile Cys Leu Gly Arg Val Gly Asp Asp Phe Trp Pro Thr Asp Pro Ser
465                 470                 475                 480
Gly Asp Ile Asn Gly Thr Tyr Trp Leu Gln Gly Cys His Met Val His
                485                 490                 495
Cys Ala Tyr Asn Ser Leu Trp Met Gly Asn Phe Ile His Pro Asp Trp
            500                 505                 510
Asp Met Phe Gln Ser Thr His Pro Cys Ala Glu Phe His Ala Ala Ser
        515                 520                 525
Arg Ala Ile Ser Gly Gly Pro Ile Tyr Val Ser Asp Val Val Gly Lys
```

530 535 540
His Asn Ile Pro Leu Leu Lys Arg Leu Val Leu Ala Asp Gly Ser Ile
545 550 555 560
Leu Arg Cys Glu Tyr His Ala Leu Pro Thr Lys Asp Cys Leu Phe Val
565 570 575
Asp Pro Leu His Asp Gly Lys Thr Met Leu Lys Ile Trp Asn Leu Asn
580 585 590
Lys Tyr Asn Gly Val Leu Gly Val Phe Asn Cys Gln Gly Gly Gly Trp
595 600 605
Ser Arg Glu Ser Arg Lys Asn Leu Cys Phe Ser Glu Tyr Ser Lys Pro
610 615 620
Ile Ser Cys Lys Thr Ser Pro Lys Asp Val Glu Trp Glu Asn Gly His
625 630 635 640
Lys Pro Phe Pro Ile Lys Gly Val Glu Cys Phe Ala Met Tyr Phe Thr
645 650 655
Lys Glu Lys Lys Leu Ile Leu Ser Gln Leu Ser Asp Thr Ile Glu Ile
660 665 670
Ser Leu Asp Pro Phe Asp Tyr Glu Leu Ile Val Val Ser Pro Met Thr
675 680 685
Ile Leu Pro Trp Glu Ser Ile Ala Phe Ala Pro Ile Gly Leu Val Asn
690 695 700
Met Leu Asn Ala Gly Gly Ala Val Lys Ser Leu Asp Ile Ser Glu Asp
705 710 715 720
Asn Glu Asp Lys Met Val Gln Val Gly Ile Lys Gly Ala Gly Glu Met
725 730 735
Met Val Tyr Ser Ser Glu Lys Pro Lys Ala Cys Arg Val Asn Gly Glu
740 745 750
Asp Met Glu Phe Glu Tyr Glu Glu Ser Met Ile Lys Val Gln Val Thr
755 760 765
Trp Asn His Asn Ser Gly Gly Phe Thr Thr Val Glu Tyr Leu Phe
770 775 780 783

<210> 4
<211> 2690
<212> DNA
<213> Beta vulgaris L.

<220>
<221> CDS
<222> (236)...(2587)

<400> 4

ctaccaaatt ccacaactta aagttcacct caatctttat tccatttttc ctccctaaac 60
ttcattgtta agattttgta attgaattca aattcttaat tctgaatttt gtcatttttt 120
ttgtggggat atttataact atcatattat ttgtgtagat cattctacaa aaaagagagt 180
gagtttttt agctcttatt tcctaagaaa ttaatagcaa aagttttgca taact atg 238
Met

gct cca agc ttt agc aag gaa aat tcc aag acg tgt gat gag gtt gca 286
Ala Pro Ser Phe Ser Lys Glu Asn Ser Lys Thr Cys Asp Glu Val Ala
5 10 15

aac cat gat gat tgc aac acg tgt cca ata att tcc ttg gaa gaa tca 334
Asn His Asp Asp Cys Asn Thr Cys Pro Ile Ile Ser Leu Glu Glu Ser
20 25 30

aac ttc atg gtg aat ggt cac gtg ata ttg tcc caa gtt cca tcc aac 382
Asn Phe Met Val Asn Gly His Val Ile Leu Ser Gln Val Pro Ser Asn
35 40 45

atc acg gcc att agt aaa atg ggt ttt gat ggg ctt ttt gtg ggt ttt 430
Ile Thr Ala Ile Ser Lys Met Gly Phe Asp Gly Leu Phe Val Gly Phe
50 55 60 65

gat gct cca gag ccc aag gcc cgg cac gtt gta tcc gtg ggc cag ctc 478
Asp Ala Pro Glu Pro Lys Ala Arg His Val Val Ser Val Gly Gln Leu
70 75 80

aag gga att ccc ttc atg agt atc ttc agg ttc aag gta tgg tgg act 526
Lys Gly Ile Pro Phe Met Ser Ile Phe Arg Phe Lys Val Trp Trp Thr
85 90 95

acc cat tgg act ggg tcc aat ggg cgg gac ctt gag cat gag acc caa 574
Thr His Trp Thr Gly Ser Asn Gly Arg Asp Leu Glu His Glu Thr Gln
100 105 110

att ctc atc ctt gat aag tca gat gaa ggt ttg ggc cgt ccc tat att 622
Ile Leu Ile Leu Asp Lys Ser Asp Glu Gly Leu Gly Arg Pro Tyr Ile
115 120 125

gtg atc ctc cca ttg atc gaa ggc cca ttt cgg gca tct ctc cag ccg 670
Val Ile Leu Pro Leu Ile Glu Gly Pro Phe Arg Ala Ser Leu Gln Pro
130 135 140 145

ggt tct gtt gat gac tat gtg gat ata tgt gtt gag agt ggg tcc act 718
Gly Ser Val Asp Asp Tyr Val Asp Ile Cys Val Glu Ser Gly Ser Thr
150 155 160

aaa gtt gtc gga gac tcg ttc cgg gct gtt ctt tat ata cgg gct ggg 766
Lys Val Val Gly Asp Ser Phe Arg Ala Val Leu Tyr Ile Arg Ala Gly
165 170 175

```
cct gac cca ttt aag tta att aaa gat aca atg aag gaa gtc caa gcc      814
Pro Asp Pro Phe Lys Leu Ile Lys Asp Thr Met Lys Glu Val Gln Ala
        180                 185                 190
cat tta ggg act ttc aaa ctc tta gat gac aaa act cct cca gga ata      862
His Leu Gly Thr Phe Lys Leu Leu Asp Asp Lys Thr Pro Pro Gly Ile
    195                 200                 205
gtg gac aag ttt gga tgg tgt aca tgg gat gca ttt tac ctc aaa gta      910
Val Asp Lys Phe Gly Trp Cys Thr Trp Asp Ala Phe Tyr Leu Lys Val
210                 215                 220                 225
gag ccw tat ggt gtt tgg gaa gga gtt aaa gga ctc gtc gaa aac ggg      958
Glu Pro Tyr Gly Val Trp Glu Gly Val Lys Gly Leu Val Glu Asn Gly
                230                 235                 240
gtc cca ccc ggt ctc gta ctc att gat gat ggg tgg caa tct att tgt     1006
Val Pro Pro Gly Leu Val Leu Ile Asp Asp Gly Trp Gln Ser Ile Cys
            245                 250                 255
cat gac gat gat ccg att acc gac caa gaa ggg ata aac cgg act tct     1054
His Asp Asp Asp Pro Ile Thr Asp Gln Glu Gly Ile Asn Arg Thr Ser
        260                 265                 270
gcc ggc gag caa atg cca tgt aga ttg atc aag tac gag gaa aac ttc     1102
Ala Gly Glu Gln Met Pro Cys Arg Leu Ile Lys Tyr Glu Glu Asn Phe
    275                 280                 285
aag ttt agg gac tat aaa agc cca aat att atg ggc cat gag gat cat     1150
Lys Phe Arg Asp Tyr Lys Ser Pro Asn Ile Met Gly His Glu Asp His
290                 295                 300                 305
ccc aat atg gga atg agg gcc ttt gtt agg gac ctt aag gag gag ttc     1198
Pro Asn Met Gly Met Arg Ala Phe Val Arg Asp Leu Lys Glu Glu Phe
                310                 315                 320
aaa act gtt gag cat gtg tat gtt tgg cat gct ttt acg ggc tat tgg     1246
Lys Thr Val Glu His Val Tyr Val Trp His Ala Phe Thr Gly Tyr Trp
            325                 330                 335
gga ggg gta agg ccc aat gtt cca ggc cta ccr gag gcc caa gta gta     1294
Gly Gly Val Arg Pro Asn Val Pro Gly Leu Pro Glu Ala Gln Val Val
        340                 345                 350
acc cca aag ctt tcc ccg ggt ctt gag atg aca atg gaa gat cta gct     1342
Thr Pro Lys Leu Ser Pro Gly Leu Glu Met Thr Met Glu Asp Leu Ala
    355                 360                 365
gtg gat aaa att gtt aat aat ggt att ggg ctt gtc cag cct gat aag     1390
Val Asp Lys Ile Val Asn Asn Gly Ile Gly Leu Val Gln Pro Asp Lys
370                 375                 380                 385
gcc caa gaa ctt tat gaa ggg ttg cat tct cat ttg gaa aat tgt ggg     1438
```

```
Ala Gln Glu Leu Tyr Glu Gly Leu His Ser His Leu Glu Asn Cys Gly
                390                 395                 400
att gat gga gtc aaa gtt gat gtc atc cat ttg ttg gag atg atg gca  1486
Ile Asp Gly Val Lys Val Asp Val Ile His Leu Leu Glu Met Met Ala
            405                 410                 415
gag gac tat gga gga aga gtt gaa cta gca aaa aca tac tat aag gca  1534
Glu Asp Tyr Gly Gly Arg Val Glu Leu Ala Lys Thr Tyr Tyr Lys Ala
        420                 425                 430
ata aca gaa tca gtg cgt aag cat ttc aaa ggc aac ggt gtg att gct  1582
Ile Thr Glu Ser Val Arg Lys His Phe Lys Gly Asn Gly Val Ile Ala
    435                 440                 445
agc atg gag cag tgc aac gat ttc atg ctc ctt ggt act gag acc att  1630
Ser Met Glu Gln Cys Asn Asp Phe Met Leu Leu Gly Thr Glu Thr Ile
450                 455                 460                 465
tgt ctt ggt cgc gtt ggg gat gac ttt tgg cca act gat ccg tct gga  1678
Cys Leu Gly Arg Val Gly Asp Asp Phe Trp Pro Thr Asp Pro Ser Gly
                470                 475                 480
gat ata aat ggt aca tat tgg ctc caa ggc tgt cat atg gtg cat tgt  1726
Asp Ile Asn Gly Thr Tyr Trp Leu Gln Gly Cys His Met Val His Cys
            485                 490                 495
gcc tac aat agc tta tgg atg gga aac ttt ata cac cct gac tgg gac  1774
Ala Tyr Asn Ser Leu Trp Met Gly Asn Phe Ile His Pro Asp Trp Asp
        500                 505                 510
atg ttc caa tct aca cac cct tgt gct gaa ttt cat gct gca tct cgt  1822
Met Phe Gln Ser Thr His Pro Cys Ala Glu Phe His Ala Ala Ser Arg
    515                 520                 525
gcg att tct ggt gga cca att tat gtt agt gat gtt gtt ggc aag cat  1870
Ala Ile Ser Gly Gly Pro Ile Tyr Val Ser Asp Val Val Gly Lys His
530                 535                 540                 545
aac atc ccc ttg ctc aaa agg ctc gtc ttg gct gat ggt tcg atc ctt  1918
Asn Ile Pro Leu Leu Lys Arg Leu Val Leu Ala Asp Gly Ser Ile Leu
                550                 555                 560
cgt tgc gag tac cat gca ctt cct act aag gat tgc cta ttt gta gat  1966
Arg Cys Glu Tyr His Ala Leu Pro Thr Lys Asp Cys Leu Phe Val Asp
            565                 570                 575
cct ttg cac gat ggc aaa aca atg ctc aaa att tgg aac ctc aac aag  2014
Pro Leu His Asp Gly Lys Thr Met Leu Lys Ile Trp Asn Leu Asn Lys
        580                 585                 590
tac aat gga gtg ctt gga gtc ttc aat tgc caa gga gga ggg tgg agc  2062
Tyr Asn Gly Val Leu Gly Val Phe Asn Cys Gln Gly Gly Gly Trp Ser
```

595 600 605

cgt gag tct cga aaa aat cta tgt ttc tca gag tat tca aaa cct att 2110
Arg Glu Ser Arg Lys Asn Leu Cys Phe Ser Glu Tyr Ser Lys Pro Ile
610 615 620 625

tcc tgc aag aca agt cca aaa gat gtt gaa tgg gag aac gga cac aag 2158
Ser Cys Lys Thr Ser Pro Lys Asp Val Glu Trp Glu Asn Gly His Lys
630 635 640

cca ttc ccc atc aaa gga gtg gaa tgt ttt gcc atg tac ttc acc aag 2206
Pro Phe Pro Ile Lys Gly Val Glu Cys Phe Ala Met Tyr Phe Thr Lys
645 650 655

gaa aaa aag cta atc ctc tca caa cta tct gac acc att gaa ata tca 2254
Glu Lys Lys Leu Ile Leu Ser Gln Leu Ser Asp Thr Ile Glu Ile Ser
660 665 670

ctt gat ccc ttc gat tac gag ctt att gta gtc tct ccg atg aca att 2302
Leu Asp Pro Phe Asp Tyr Glu Leu Ile Val Val Ser Pro Met Thr Ile
675 680 685

cta ccc tgg gag tcg atc gca ttt gca ccc ata gga tta gta aac atg 2350
Leu Pro Trp Glu Ser Ile Ala Phe Ala Pro Ile Gly Leu Val Asn Met
690 695 700 705

ctc aac gcc gga ggg gca gtc aag tct ttg gac atc agt gag gat aat 2398
Leu Asn Ala Gly Gly Ala Val Lys Ser Leu Asp Ile Ser Glu Asp Asn
710 715 720

gag gat aag atg gtt cag gtt ggt att aaa ggg gcc gga gaa atg atg 2446
Glu Asp Lys Met Val Gln Val Gly Ile Lys Gly Ala Gly Glu Met Met
725 730 735

gtt tat tca tca gaa aag cca aaa gcg tgt aga gtt aat gga gaa gac 2494
Val Tyr Ser Ser Glu Lys Pro Lys Ala Cys Arg Val Asn Gly Glu Asp
740 745 750

atg gag ttt gag tat gaa gag agc atg att aag gtt caa gtt aca tgg 2542
Met Glu Phe Glu Tyr Glu Glu Ser Met Ile Lys Val Gln Val Thr Trp
755 760 765

aac cat aac tca ggt ggt ttt acc act gtt gag tac tta ttt tga gcttg 2592
Asn His Asn Ser Gly Gly Phe Thr Thr Val Glu Tyr Leu Phe
770 775 780

aagctaatct aagtctttac ttaatgagtg atgtaactga gtagttgact tgagagtaca 2652

gtatgtgtga agcttattat tccaaaaaaa aaaaaaaa 2690

<210> 5
<211> 777
<212> PRT

<213> Brassica juncea

<400> 5

```
Met Ala Pro Pro Ser Val Ile Lys Ser Asp Ala Ala Val Asn Gly Ile
 1               5                  10                  15
Asp Leu Ser Gly Lys Pro Leu Phe Arg Leu Glu Gly Ser Asp Leu Leu
            20                  25                  30
Ala Asn Gly His Val Val Leu Thr Asp Val Pro Val Asn Val Thr Val
        35                  40                  45
Thr Ala Ser Pro Tyr Leu Ala Asp Lys Asp Gly Glu Pro Val Asp Ala
    50                  55                  60
Ser Ala Gly Ser Phe Ile Gly Phe Asn Leu Asp Gly Glu Pro Arg Ser
65                  70                  75                  80
Arg His Val Ala Ser Ile Gly Lys Leu Arg Asp Ile Arg Phe Met Ser
                85                  90                  95
Ile Phe Arg Phe Lys Val Trp Trp Thr Thr His Trp Val Gly Ser Lys
            100                 105                 110
Gly Ser Asp Ile Glu Asn Glu Thr Gln Ile Ile Ile Leu Glu Asn Ser
        115                 120                 125
Gly Ser Gly Arg Pro Tyr Val Leu Leu Leu Pro Leu Leu Glu Gly Ser
    130                 135                 140
Phe Arg Ser Ser Phe Gln Pro Gly Glu Asp Asp Asp Val Ala Val Cys
145                 150                 155                 160
Val Glu Ser Gly Ser Thr Gln Val Thr Gly Ser Glu Phe Arg Gln Val
                165                 170                 175
Val Tyr Val His Ala Gly Asp Asp Pro Phe Lys Leu Val Lys Asp Ala
            180                 185                 190
Met Lys Val Val Arg Val His Met Asn Thr Phe Lys Leu Leu Glu Glu
        195                 200                 205
Lys Thr Pro Pro Gly Ile Val Asp Lys Phe Gly Trp Cys Thr Trp Asp
    210                 215                 220
Ala Phe Tyr Leu Thr Val Asn Pro Asp Gly Val His Lys Gly Val Lys
225                 230                 235                 240
Cys Leu Val Asp Gly Gly Cys Pro Pro Gly Leu Val Leu Ile Asp Asp
                245                 250                 255
Gly Trp Gln Ser Ile Gly His Asp Ser Asp Gly Ile Asp Val Glu Gly
            260                 265                 270
Met Ser Cys Thr Val Ala Gly Glu Gln Met Pro Cys Arg Leu Leu Lys
        275                 280                 285
Phe Gln Glu Asn Phe Lys Phe Arg Asp Tyr Val Ser Pro Lys Asp Lys
```

290 295 300
Asn Glu Val Gly Met Lys Ala Phe Val Arg Asp Leu Lys Glu Glu Phe
305 310 315 320
Ser Thr Val Asp Tyr Ile Tyr Val Trp His Ala Leu Cys Gly Tyr Trp
325 330 335
Gly Gly Leu Arg Pro Gly Ala Pro Thr Leu Pro Pro Ser Thr Ile Val
340 345 350
Arg Pro Glu Leu Ser Pro Gly Leu Lys Leu Thr Met Gln Asp Leu Ala
355 360 365
Val Asp Lys Ile Val Asp Thr Gly Ile Gly Phe Val Ser Pro Asp Met
370 375 380
Ala Asn Glu Phe Tyr Glu Gly Leu His Ser His Leu Gln Asn Val Gly
385 390 395 400
Ile Asp Gly Val Lys Val Asp Val Ile His Ile Leu Glu Met Leu Cys
405 410 415
Glu Lys Tyr Gly Gly Arg Val Asp Leu Ala Lys Ala Tyr Phe Lys Ala
420 425 430
Leu Thr Ser Ser Val Asn Lys His Phe Asp Gly Asn Gly Val Ile Ala
435 440 445
Ser Met Glu His Cys Asn Asp Phe Met Phe Leu Gly Thr Glu Ala Ile
450 455 460
Ser Leu Gly Arg Val Gly Asp Asp Phe Trp Cys Thr Asp Pro Ser Gly
465 470 475 480
Asp Ile Asn Gly Thr Tyr Trp Leu Gln Gly Cys His Met Val His Cys
485 490 495
Ala Tyr Asn Ser Leu Trp Met Gly Asn Phe Ile Gln Pro Asp Trp Asp
500 505 510
Met Phe Gln Ser Thr His Pro Cys Ala Glu Phe His Ala Ala Ser Arg
515 520 525
Ala Ile Ser Gly Gly Pro Ile Tyr Ile Ser Asp Cys Val Gly Gln His
530 535 540
Asp Phe Asp Leu Leu Lys Arg Leu Val Leu Pro Asp Gly Ser Ile Leu
545 550 555 560
Arg Cys Glu His Tyr Ala Leu Pro Thr Arg Asp Arg Leu Phe Glu Asp
565 570 575
Pro Leu His Asp Gly Lys Thr Met Leu Lys Ile Trp Asn Leu Asn Lys
580 585 590
Tyr Thr Gly Ile Ile Gly Ala Phe Asn Cys Gln Gly Gly Gly Trp Cys
595 600 605
Arg Glu Thr Arg Arg Asn Gln Cys Phe Ser Gln Cys Val Asn Thr Leu

610 615 620
Thr Ala Thr Thr Asn Pro Lys Asp Val Glu Trp Asn Ser Gly Asn Asn
625 630 635 640
Pro Ile Ser Val Glu Asn Val Glu Glu Phe Ala Leu Phe Leu Ser Gln
645 650 655
Ser Lys Lys Leu Val Leu Ser Gly Pro Asn Asp Asp Leu Glu Ile Thr
660 665 670
Leu Glu Pro Phe Lys Phe Glu Leu Ile Thr Val Ser Pro Val Val Thr
675 680 685
Ile Glu Gly Ser Ser Val Gln Phe Ala Pro Ile Gly Leu Val Asn Met
690 695 700
Leu Asn Thr Ser Gly Ala Ile Arg Ser Leu Val Tyr His Glu Glu Ser
705 710 715 720
Val Glu Ile Gly Val Arg Gly Ala Gly Glu Phe Arg Val Tyr Ala Ser
725 730 735
Arg Lys Pro Ala Ser Cys Lys Ile Asp Gly Glu Val Val Glu Phe Gly
740 745 750
Tyr Glu Glu Ser Met Val Met Val Gln Val Pro Trp Ser Ala Pro Glu
755 760 765
Gly Leu Ser Ser Ile Lys Tyr Glu Phe
770 775 777

<210> 6
<211> 2690
<212> DNA
<213> Brassica juncea

<220>
<221> CDS
<222> (134)...(2467)

<400> 6
accaatccaa aatctcatca aataatcgca attaggggaa gtttacaaga ttcatcatct 60
ccgttactat ataactacgc tcttcttcct tcgcctaatc caacttaacc taaaaaccac 120
tctatcagcg aaa atg gct cca ccg agc gta att aaa tcc gat gct gca 169
Met Ala Pro Pro Ser Val Ile Lys Ser Asp Ala Ala
5 10
gtc aac ggc att gac ctc tcc gga aag ccg ctt ttc cgg cta gag ggt 217
Val Asn Gly Ile Asp Leu Ser Gly Lys Pro Leu Phe Arg Leu Glu Gly
15 20 25

```
tcc gat ctc cta gcc aat ggt cac gtt gtc tta acc gat gta ccg gtt     265
Ser Asp Leu Leu Ala Asn Gly His Val Val Leu Thr Asp Val Pro Val
     30                  35                  40
aac gtg act gtc act gct tca cct tac cta gct gac aaa gac gga gaa     313
Asn Val Thr Val Thr Ala Ser Pro Tyr Leu Ala Asp Lys Asp Gly Glu
 45                  50                  55                  60
ccg gtt gac gcc tcc gct ggt tca ttc atc ggg ttt aat ctc gac ggt     361
Pro Val Asp Ala Ser Ala Gly Ser Phe Ile Gly Phe Asn Leu Asp Gly
                 65                  70                  75
gag cca cga agc cgc cac gtg gcg tcc atc ggt aaa ctc agg gat att     409
Glu Pro Arg Ser Arg His Val Ala Ser Ile Gly Lys Leu Arg Asp Ile
             80                  85                  90
cga ttc atg agc ata ttc cgt ttc aag gtt tgg tgg act act cac tgg     457
Arg Phe Met Ser Ile Phe Arg Phe Lys Val Trp Trp Thr Thr His Trp
         95                  100                 105
gtc ggt tcc aaa gga tcc gac atc gag aac gag acc cag atc atc atc     505
Val Gly Ser Lys Gly Ser Asp Ile Glu Asn Glu Thr Gln Ile Ile Ile
    110                 115                 120
ctc gag aac tcc ggg tcg ggt cgt cct tat gtt ctt ctt ctg ccg ctt     553
Leu Glu Asn Ser Gly Ser Gly Arg Pro Tyr Val Leu Leu Leu Pro Leu
125                 130                 135                 140
ctt gaa ggc tct ttc cgt tca tcc ttt cag cct ggg gaa gac gat gac     601
Leu Glu Gly Ser Phe Arg Ser Ser Phe Gln Pro Gly Glu Asp Asp Asp
                145                 150                 155
gtg gcg gtt tgt gtc gaa tcc ggg tcg acc cag gtg acc ggg tcg gag     649
Val Ala Val Cys Val Glu Ser Gly Ser Thr Gln Val Thr Gly Ser Glu
            160                 165                 170
ttt cgt caa gtt gtg tat gtt cac gcc gga gac gat ccg ttc aag ctc     697
Phe Arg Gln Val Val Tyr Val His Ala Gly Asp Asp Pro Phe Lys Leu
        175                 180                 185
gtg aaa gac gcg atg aag gtg gtt agg gtt cat atg aac acc ttc aag     745
Val Lys Asp Ala Met Lys Val Val Arg Val His Met Asn Thr Phe Lys
    190                 195                 200
ctc ttg gaa gag aag acr ccg ccg gga atc gtc gat aag ttc ggg tgg     793
Leu Leu Glu Glu Lys Thr Pro Pro Gly Ile Val Asp Lys Phe Gly Trp
205                 210                 215                 220
tgc acg tgg gat gcg ttt tat ttg acg gtg aac cct gac gga gtt cat     841
Cys Thr Trp Asp Ala Phe Tyr Leu Thr Val Asn Pro Asp Gly Val His
                225                 230                 235
aag ggt gtt aag tgt ctc gtc gac ggt ggt tgt ccg ccg gga ttg gtc     889
```

```
Lys Gly Val Lys Cys Leu Val Asp Gly Gly Cys Pro Pro Gly Leu Val
            240                 245                 250
cta atc gac gac ggt tgg caa tcg att gga cat gac tcc gat ggt atc     937
Leu Ile Asp Asp Gly Trp Gln Ser Ile Gly His Asp Ser Asp Gly Ile
        255                 260                 265
gat gtt gaa ggg atg agt tgt acc gtc gcc ggg gag caa atg cct tgc     985
Asp Val Glu Gly Met Ser Cys Thr Val Ala Gly Glu Gln Met Pro Cys
    270                 275                 280
agg ctt ctg aaa ttt caa gag aac ttc aag ttc aga gac tac gtc tct    1033
Arg Leu Leu Lys Phe Gln Glu Asn Phe Lys Phe Arg Asp Tyr Val Ser
285                 290                 295                 300
ccg aaa gac aaa aac gaa gtc ggg atg aaa gct ttc gtc aga gat ctg    1081
Pro Lys Asp Lys Asn Glu Val Gly Met Lys Ala Phe Val Arg Asp Leu
                305                 310                 315
aaa gaa gaa ttc tcc acc gtt gat tac atc tac gtc tgg cac gcg ctt    1129
Lys Glu Glu Phe Ser Thr Val Asp Tyr Ile Tyr Val Trp His Ala Leu
            320                 325                 330
tgc ggc tac tgg ggt ggt ctt cgt ccc gga gct cct act ctt ccg ccc    1177
Cys Gly Tyr Trp Gly Gly Leu Arg Pro Gly Ala Pro Thr Leu Pro Pro
        335                 340                 345
tca act att gtc cgg cca gag ctc tcg ccg ggg ctt aag ttg acg atg    1225
Ser Thr Ile Val Arg Pro Glu Leu Ser Pro Gly Leu Lys Leu Thr Met
    350                 355                 360
caa gat ctc gcc gtt gat aag att gtc gat acc gga atc gga ttc gtc    1273
Gln Asp Leu Ala Val Asp Lys Ile Val Asp Thr Gly Ile Gly Phe Val
365                 370                 375                 380
tcg ccg gac atg gcg aat gag ttt tac gaa ggt ctt cac tct cat ctt    1321
Ser Pro Asp Met Ala Asn Glu Phe Tyr Glu Gly Leu His Ser His Leu
                385                 390                 395
caa aac gtc ggt att gac ggc gtt aaa gtt gac gtc atc cac ata ttg    1369
Gln Asn Val Gly Ile Asp Gly Val Lys Val Asp Val Ile His Ile Leu
            400                 405                 410
gag atg ttg tgc gag aaa tat ggc ggg aga gta gac ttg gct aaa gct    1417
Glu Met Leu Cys Glu Lys Tyr Gly Gly Arg Val Asp Leu Ala Lys Ala
        415                 420                 425
tac ttc aag gcg tta act tcc tca gtg aat aag cat ttt gac ggt aac    1465
Tyr Phe Lys Ala Leu Thr Ser Ser Val Asn Lys His Phe Asp Gly Asn
    430                 435                 440
ggc gtt atc gct agc atg gag cac tgt aat gat ttc atg ttc ctt gga    1513
Gly Val Ile Ala Ser Met Glu His Cys Asn Asp Phe Met Phe Leu Gly
```

```
445                 450                 455                 460
acc gaa gcc atc tct cta ggt cgt gtc ggt gat gac ttt tgg tgc acg    1561
Thr Glu Ala Ile Ser Leu Gly Arg Val Gly Asp Asp Phe Trp Cys Thr
                465                 470                 475
gat cca tca ggc gac ata aac ggc aca tat tgg ctg caa gga tgc cac    1609
Asp Pro Ser Gly Asp Ile Asn Gly Thr Tyr Trp Leu Gln Gly Cys His
            480                 485                 490
atg gtc cac tgt gcc tac aac agt ctt tgg atg gga aat ttc atc cag    1657
Met Val His Cys Ala Tyr Asn Ser Leu Trp Met Gly Asn Phe Ile Gln
        495                 500                 505
cct gat tgg gac atg ttt cag tcc aca cat cct tgt gct gag ttc cat    1705
Pro Asp Trp Asp Met Phe Gln Ser Thr His Pro Cys Ala Glu Phe His
    510                 515                 520
gct gct tct cgt gcc atc tcc ggt ggg ccc att tac atc agc gat tgt    1753
Ala Ala Ser Arg Ala Ile Ser Gly Gly Pro Ile Tyr Ile Ser Asp Cys
525                 530                 535                 540
gtg ggc cag cac gat ttc gat ctc ttg aag cga ctc gtc ttg cct gac    1801
Val Gly Gln His Asp Phe Asp Leu Leu Lys Arg Leu Val Leu Pro Asp
                545                 550                 555
ggt tcg att ttg agg tgt gag cac tat gca ctc cca act cgt gac cgt    1849
Gly Ser Ile Leu Arg Cys Glu His Tyr Ala Leu Pro Thr Arg Asp Arg
            560                 565                 570
ctc ttt gaa gac cct ctt cat gat ggc aaa acc atg ctc aag att tgg    1897
Leu Phe Glu Asp Pro Leu His Asp Gly Lys Thr Met Leu Lys Ile Trp
        575                 580                 585
aac ttg aac aag tac act gga att att gga gca ttc aac tgc caa gga    1945
Asn Leu Asn Lys Tyr Thr Gly Ile Ile Gly Ala Phe Asn Cys Gln Gly
    590                 595                 600
gga gga tgg tgc aga gaa acc cga cgc aac caa tgc ttc tcc caa tgc    1993
Gly Gly Trp Cys Arg Glu Thr Arg Arg Asn Gln Cys Phe Ser Gln Cys
605                 610                 615                 620
gtt aac acg tta acc gcc aca aca aat cct aag gac gtt gaa tgg aac    2041
Val Asn Thr Leu Thr Ala Thr Thr Asn Pro Lys Asp Val Glu Trp Asn
                625                 630                 635
agt ggg aac aac cca atc tcc gtt gaa aac gtt gaa gag ttt gct ttg    2089
Ser Gly Asn Asn Pro Ile Ser Val Glu Asn Val Glu Glu Phe Ala Leu
            640                 645                 650
ttc ttg tct cag tct aag aag ctt gtg ttg tct gga cca aac gat gat    2137
Phe Leu Ser Gln Ser Lys Lys Leu Val Leu Ser Gly Pro Asn Asp Asp
        655                 660                 665
```

ctc gag atc act ttg gag cct ttc aag ttt gag cta atc act gtc tca 2185
Leu Glu Ile Thr Leu Glu Pro Phe Lys Phe Glu Leu Ile Thr Val Ser
670 675 680
cca gtt gtc act att gag ggt agt tcg gtt cag ttt gct cca atc gga 2233
Pro Val Val Thr Ile Glu Gly Ser Ser Val Gln Phe Ala Pro Ile Gly
685 690 695 700
ttg gtt aac atg cta aac act agc ggt gca att cga tcc ttg gtg tat 2281
Leu Val Asn Met Leu Asn Thr Ser Gly Ala Ile Arg Ser Leu Val Tyr
705 710 715
cat gag gaa tcc gtt gag att gga gtt cgt ggt gct gga gag ttc agg 2329
His Glu Glu Ser Val Glu Ile Gly Val Arg Gly Ala Gly Glu Phe Arg
720 725 730
gtt tat gca tca agg aaa cct gcg agc tgc aaa att gat ggt gaa gtt 2377
Val Tyr Ala Ser Arg Lys Pro Ala Ser Cys Lys Ile Asp Gly Glu Val
735 740 745
gtt gag ttt gga tac gaa gag tca atg gtg atg gtt caa gtg cct tgg 2425
Val Glu Phe Gly Tyr Glu Glu Ser Met Val Met Val Gln Val Pro Trp
750 755 760
tct gca ccc gag ggt ttg tct tct att aag tat gag ttt tag agtttccga 2476
Ser Ala Pro Glu Gly Leu Ser Ser Ile Lys Tyr Glu Phe
765 770 775
aggtgcttat ttgtatcctt ctaaactcct taattatgag ctccgtgccg tttctttttc 2536
tatatggttt ctgagagtga acatctaata tttacccact agggtataat tattggcttt 2596
taagtgattt gtttttgaac tgtttttagt ggtgtaattt gtactgcccc tattattttt 2656
catatttatt tgtgaaagat aaaaaaaaaa aaaa 2690

<210> 7
<211> 572
<212> PRT
<213> Brassica napus

<400> 7
Leu Glu Glu Lys Thr Pro Pro Gly Ile Val Asp Lys Phe Gly Trp Cys
5 10 15
Thr Trp Asp Ala Phe Tyr Leu Thr Val Asn Pro Asp Gly Val His Lys
20 25 30
Gly Val Lys Cys Leu Val Asp Gly Gly Cys Pro Pro Gly Leu Val Leu
35 40 45
Ile Asp Asp Gly Trp Gln Ser Ile Gly His Asp Ser Asp Gly Ile Asp
50 55 60

Val Glu Gly Met Ser Cys Thr Val Ala Gly Glu Gln Met Pro Cys Arg
65 70 75 80
Leu Pro Lys Phe Gln Glu Asn Phe Lys Phe Arg Asp Tyr Val Ser Pro
85 90 95
Lys Asp Lys Asn Glu Val Gly Met Lys Ala Phe Val Arg Asp Leu Lys
100 105 110
Glu Glu Phe Ser Thr Val Asp Tyr Ile Tyr Val Trp His Ala Leu Cys
115 120 125
Gly Tyr Trp Gly Gly Leu Arg Pro Gly Ala Pro Thr Leu Pro Pro Ser
130 135 140
Thr Ile Val Arg Pro Glu Leu Ser Pro Gly Leu Lys Leu Thr Met Gln
145 150 155 160
Asp Leu Ala Val Asp Lys Ile Ile Asp Thr Gly Ile Gly Phe Val Ser
165 170 175
Pro Asp Met Ala Asn Glu Phe Tyr Glu Gly Leu His Ser His Leu Gln
180 185 190
Asn Val Gly Ile Asn Gly Val Lys Val Asp Val Ile His Ile Leu Glu
195 200 205
Met Leu Cys Glu Lys Tyr Gly Gly Arg Val Asp Leu Ala Lys Ala Tyr
210 215 220
Phe Lys Ala Leu Thr Ser Ser Val Asn Lys His Phe Asp Gly Asn Ala
225 230 235 240
Val Ile Ala Ser Met Glu His Cys Asn Asp Phe Met Phe Leu Gly Thr
245 250 255
Glu Ala Ile Ser Leu Gly Arg Val Gly Asp Asp Phe Trp Cys Thr Asp
260 265 270
Pro Ser Gly Asp Ile Asn Gly Thr Tyr Trp Leu Gln Gly Cys His Met
275 280 285
Val His Cys Ala Tyr Asn Ser Leu Trp Met Gly Asn Phe Ile Gln Pro
290 295 300
Asp Trp Asp Met Phe Gln Ser Thr His Pro Cys Ala Glu Phe His Ala
305 310 315 320
Ala Ser Arg Ala Ile Ser Gly Gly Pro Ile Tyr Ile Ser Asp Cys Val
325 330 335
Gly Gln His Asp Phe Asp Leu Leu Arg Arg Leu Val Leu Pro Asp Gly
340 345 350
Ser Ile Leu Arg Cys Glu Tyr Tyr Ala Leu Pro Thr Arg Asp Arg Leu
355 360 365
Phe Glu Asp Pro Leu His Asp Gly Lys Thr Met Leu Lys Ile Trp Asn
370 375 380

Leu Asn Lys Tyr Thr Gly Ile Ile Gly Ala Phe Asn Cys Gln Gly Gly
385 390 395 400
Gly Trp Cys Arg Glu Thr Arg Arg Asp Gln Cys Phe Ser Gln Cys Val
405 410 415
Asn Thr Leu Thr Ala Thr Thr Asn Pro Asn Asp Val Glu Trp Asn Ser
420 425 430
Gly Asn Asn Pro Ile Ser Ile Glu Asn Val Glu Glu Phe Ala Leu Phe
435 440 445
Leu Ser Gln Ser Lys Lys Leu Val Leu Ser Gly Gln Asn Asp Asp Leu
450 455 460
Glu Ile Thr Leu Glu Pro Phe Lys Phe Glu Leu Ile Thr Val Ser Pro
465 470 475 480
Val Val Thr Ile Glu Gly Ser Ser Val Gln Phe Ala Pro Ile Gly Leu
485 490 495
Val Asn Met Leu Asn Thr Ser Gly Ala Ile Arg Ser Leu Val Tyr His
500 505 510
Glu Glu Ser Val Glu Ile Gly Val Arg Gly Ala Gly Glu Phe Arg Val
515 520 525
Tyr Ala Ser Lys Lys Pro Val Ser Cys Lys Ile Asp Gly Glu Asp Val
530 535 540
Glu Phe Gly Tyr Glu Glu Ser Met Val Met Val Gln Val Pro Trp Ser
545 550 555 560
Ala Pro Glu Gly Leu Ser Ser Ile Lys Tyr Leu Phe
565 570 572

<210> 8
<211> 1762
<212> DNA
<213> Brassica napus

<220>
<221> CDS
<222> (1)...(1719)

<400> 8
ttg gaa gaa aaa acg ccg ccg gga atc gtc gat aag ttc ggg tgg tgc 48
Leu Glu Glu Lys Thr Pro Pro Gly Ile Val Asp Lys Phe Gly Trp Cys
5 10 15
acg tgg gat gcg ttt tat ttg acg gtg aac cct gac gga gtt cat aag 96
Thr Trp Asp Ala Phe Tyr Leu Thr Val Asn Pro Asp Gly Val His Lys

```
                    20                  25                  30
ggt gtt aag tgt ctc gtc gac ggt ggt tgt ccg ccg gga ttg gtc cta   144
Gly Val Lys Cys Leu Val Asp Gly Gly Cys Pro Pro Gly Leu Val Leu
          35                  40                  45
atc gac gac ggt tgg caa tcg att gga cat gac tcc gat ggt atc gat   192
Ile Asp Asp Gly Trp Gln Ser Ile Gly His Asp Ser Asp Gly Ile Asp
     50                  55                  60
gtt gaa ggg atg agt tgt acc gtc gcc ggg gag caa atg cct tgc agg   240
Val Glu Gly Met Ser Cys Thr Val Ala Gly Glu Gln Met Pro Cys Arg
 65                  70                  75                  80
ctt ccg aaa ttt caa gag aac ttc aag ttc aga gac tac gtc tct ccg   288
Leu Pro Lys Phe Gln Glu Asn Phe Lys Phe Arg Asp Tyr Val Ser Pro
                85                  90                  95
aaa gac aaa aac gaa gtc ggg atg aaa gct ttc gtc aga gat ctg aaa   336
Lys Asp Lys Asn Glu Val Gly Met Lys Ala Phe Val Arg Asp Leu Lys
           100                 105                 110
gaa gaa ttc tcc acc gtt gat tac atc tac gtc tgg cac gcg ctt tgc   384
Glu Glu Phe Ser Thr Val Asp Tyr Ile Tyr Val Trp His Ala Leu Cys
       115                 120                 125
ggy tac tgg ggw ggt ctt cgt ccc gga gct cct act ctt ccg ccs tcr   432
Gly Tyr Trp Gly Gly Leu Arg Pro Gly Ala Pro Thr Leu Pro Pro Ser
   130                 135                 140
act att gtc cgr cca gag ctc tcg ccg ggg ctt aag ttg acg atg caa   480
Thr Ile Val Arg Pro Glu Leu Ser Pro Gly Leu Lys Leu Thr Met Gln
145                 150                 155                 160
gat ctc gcc gtt gat aag atc atc gat acc gga atc gga ttc gtc tcg   528
Asp Leu Ala Val Asp Lys Ile Ile Asp Thr Gly Ile Gly Phe Val Ser
               165                 170                 175
ccg gac atg gcg aac gag ttt tac gaa ggt ctt cac tct cat ctt caa   576
Pro Asp Met Ala Asn Glu Phe Tyr Glu Gly Leu His Ser His Leu Gln
           180                 185                 190
aac gtc ggc att aac ggc gtt aaa gtt gac gtt atc cac ata ctg gag   624
Asn Val Gly Ile Asn Gly Val Lys Val Asp Val Ile His Ile Leu Glu
       195                 200                 205
atg ttg tgc gag aaa tat ggc ggg aga gtt gac ttg gct aaa gct tac   672
Met Leu Cys Glu Lys Tyr Gly Gly Arg Val Asp Leu Ala Lys Ala Tyr
   210                 215                 220
ttc aag gcg tta acg tcg tca gtg aat aag cat ttt gac ggc aac gcc   720
Phe Lys Ala Leu Thr Ser Ser Val Asn Lys His Phe Asp Gly Asn Ala
225                 230                 235                 240
```

```
gtt atc gcc agc atg gag cac tgt aat gac ttc atg ttc ctt gga acc     768
Val Ile Ala Ser Met Glu His Cys Asn Asp Phe Met Phe Leu Gly Thr
                245                 250                 255
gaa gcc atc tct cta ggt cgt gtc ggt gat gac ttt tgg tgc acg gat     816
Glu Ala Ile Ser Leu Gly Arg Val Gly Asp Asp Phe Trp Cys Thr Asp
            260                 265                 270
cca tct ggc gac att aac ggc acg tat tgg ctg caa gga tgt cac atg     864
Pro Ser Gly Asp Ile Asn Gly Thr Tyr Trp Leu Gln Gly Cys His Met
        275                 280                 285
gtc cac tgt gcc tac aac agt ctt tgg atg gga aat ttc atc cag cct     912
Val His Cys Ala Tyr Asn Ser Leu Trp Met Gly Asn Phe Ile Gln Pro
    290                 295                 300
gat tgg gac atg ttt cag tcc aca cat cct tgt gct gag ttc cat gct     960
Asp Trp Asp Met Phe Gln Ser Thr His Pro Cys Ala Glu Phe His Ala
305                 310                 315                 320
gct tca cgt gcc atc tcc ggt ggg ccc att tac atc agc gat tgt gtg    1008
Ala Ser Arg Ala Ile Ser Gly Gly Pro Ile Tyr Ile Ser Asp Cys Val
                325                 330                 335
ggc cag cac gat ttc gat ctc ttg agg aga ctc gtt ttg cct gac ggt    1056
Gly Gln His Asp Phe Asp Leu Leu Arg Arg Leu Val Leu Pro Asp Gly
            340                 345                 350
tcg att ttg agg tgt gag tac tat gct ctc cca act cgt gac cgt ctc    1104
Ser Ile Leu Arg Cys Glu Tyr Tyr Ala Leu Pro Thr Arg Asp Arg Leu
        355                 360                 365
ttt gaa gac cct ctt cat gat ggc aaa acc atg ctc aag att tgg aac    1152
Phe Glu Asp Pro Leu His Asp Gly Lys Thr Met Leu Lys Ile Trp Asn
    370                 375                 380
ttg aac aag tac act gga atc atc gga gca ttc aac tgt caa gga gga    1200
Leu Asn Lys Tyr Thr Gly Ile Ile Gly Ala Phe Asn Cys Gln Gly Gly
385                 390                 395                 400
gga tgg tgc aga gaa act cga cgc gac caa tgc ttc tcc caa tgc gtt    1248
Gly Trp Cys Arg Glu Thr Arg Arg Asp Gln Cys Phe Ser Gln Cys Val
                405                 410                 415
aac acg tta acc gcc aca aca aat cct aat gac gtt gaa tgg aac agt    1296
Asn Thr Leu Thr Ala Thr Thr Asn Pro Asn Asp Val Glu Trp Asn Ser
            420                 425                 430
ggg aac aac ccg atc tcc att gaa aac gtt gaa gag ttt gct ttg ttc    1344
Gly Asn Asn Pro Ile Ser Ile Glu Asn Val Glu Glu Phe Ala Leu Phe
        435                 440                 445
ttg tct caa tcc aag aag ctt gtg ttg tcc ggg caa aac gat gat ctc    1392
```

Leu Ser Gln Ser Lys Lys Leu Val Leu Ser Gly Gln Asn Asp Asp Leu
450 455 460

gag atc aca tta gag ccc ttc aag ttc gag ctc atc act gtc tca cca 1440
Glu Ile Thr Leu Glu Pro Phe Lys Phe Glu Leu Ile Thr Val Ser Pro
465 470 475 480

gtt gtc acc att gag ggc agt tcg gtt cag ttt gct cca atc gga ttg 1488
Val Val Thr Ile Glu Gly Ser Ser Val Gln Phe Ala Pro Ile Gly Leu
485 490 495

gtt aac atg ctt aac act agc ggt gcg att cga tcc ttg gtt tat cat 1536
Val Asn Met Leu Asn Thr Ser Gly Ala Ile Arg Ser Leu Val Tyr His
500 505 510

gag gaa tcc gtt gag atc ggt gtt cgt ggt gct gga gaa ttc agg gtt 1584
Glu Glu Ser Val Glu Ile Gly Val Arg Gly Ala Gly Glu Phe Arg Val
515 520 525

tat gca tcg aag aaa cct gtg agc tgc aag att gat ggt gaa gat gtt 1632
Tyr Ala Ser Lys Lys Pro Val Ser Cys Lys Ile Asp Gly Glu Asp Val
530 535 540

gag ttt ggg tac gaa gag tca atg gtg atg gtt caa gtg cct tgg tct 1680
Glu Phe Gly Tyr Glu Glu Ser Met Val Met Val Gln Val Pro Trp Ser
545 550 555 560

gca cca gag ggt ttg tct tct att aag tat ttg ttt tag agttatttaa 1729
Ala Pro Glu Gly Leu Ser Ser Ile Lys Tyr Leu Phe
565 570

ggtgcttaat tgaaaaaaaa aaaaaaaaaa aaa 1762

<210> 9
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Designed oligonucleotide primer to obtain raffinose synthase gene.

<400> 9
ccaatctgat catgcttgtg ccgaa 25

<210> 10
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Designed oligonucleotide primer to obtain raffinose synthase gene.

<400> 10
ggaacaaagt tatgcactat tatttaaggt 30

<210> 11
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Designed oligonucleotide primer to obtain raffinose synthase gene.

<400> 11
ctaccaaatt ccacaactta aagttca 27

<210> 12
<211> 32
<212> DNA
<213> Artificial Sequence

<220>
<223> Designed oligonucleotide primer to obtain raffinose synthase gene.

<400> 12
ggaataataa gcttcacaca tactgtactc tc 32

<210> 13
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Designed oligonucleotide primer to obatin raffinose synthase gene.

<400> 13
atggctccaa gctttagcaa ggaaaattcc 30

<210> 14
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Designed oligonucleotide primer to obtain raffinose synthase gene.

<400> 14
tcaaaataag tactcaacag tggtaaaacc 30

<210> 15
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Designed oligonucleotide primer to obtain raffinose synthase gene.

<400> 15
ttggaagaga agacgccgcc gggaatcgtc 30

<210> 16
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Designed oligonucleotide primer to obtain raffinose synthase gene.

<400> 16
ttaagccccg gcgagagctc tggccggaca 30

<210> 17
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Designed oligonucleotide primer to obtain raffinose synthase gene.

<400> 17
accaatccaa aatctcatca aataatcgca 30

<210> 18
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Designed oligonucleotide primer to obatin raffinose synthase gene.

<400> 18
aaataatagg ggcagtacaa attacaccac 30

<210> 19
<211> 29
<212> DNA
<213> Artificial Sequence

<220>
<223> Designed oligonucleotide primer to obtain raffinose synthase gene.

<400> 19
atggctccac cgagcgtaat taaatccga 29

<210> 20
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Designed oligonucleotide primer to obtain raffinose synthase gene.

<400> 20
ctaaaactca tacttaatag aagacaaacc 30

<210> 21
<211> 41
<212> DNA

<213> Artificial Sequence

<220>
<223> Designed oligonucleotide primer to obtain raffinose synthase gene, n is i.

<400> 21
cgatggatgg gnaanttnat ncanccngan tggganatgt t 41

<210> 22
<211> 32
<212> DNA
<213> Artificial Sequence

<220>
<223> Designed oligonucleotide primer to obtain raffinose synthase gene, n is i, r is a or g.

<400> 22
ggccacatnt tnacnarncc natnggngcn aa 32

<210> 23
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Designed oligonucleotide primer to obtain raffinose synthase gene.

<400> 23
tgttactagg cgaaacaaga gtagctctga 30

<210> 24
<211> 47
<212> DNA
<213> Artificial Sequence

<220>
<223> Designed oligonucleotide primer to obtain raffinose synthase gene, n is i.

<400> 24
cgaggnggnt gnccnccngg nttngtnatn atngangang gntggca 47

<210> 25
<211> 29
<212> DNA
<213> Artificial Sequence

<220>
<223> Designed oligonucleotide primer to obtain raffinose synthase gene, n is i, y is t or c, r is a or g.

<400> 25
atyttrtcna cngcnarrtc ytccatngt 29

<210> 26
<211> 38
<212> DNA
<213> Artificial Sequence

<220>
<223> Designed oligonucleotide primer to obtain raffinose synthase gene, n is i, y is t or c.

<400> 26
ggnacntant ggytncangg ntgncanatg gtncantg 38

<210> 27
<211> 32
<212> DNA
<213> Artificial Sequence

<220>
<223> Designed oligonucleotide primer to obtain raffinose synthase gene, n is i, r is a or g.

<400> 27
ggccacatnt tnacnarncc natnggngcn aa 32

<210> 28
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Designed oligonucleotide primer to obtain raffinose synthase gene.

<400> 28
atctatttgt catgacgatg atccga 26

<210> 29
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Designed oligonucleotide primer to obtain raffinose synthase gene.

<400> 29
ggccctcatt cccatattgg gatgatcctc 30

<210> 30
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Designed oligonucleotide primer to obtain raffinose synthase gene.

<400> 30

aagcatgcca aacatacaca tgctcaacag 30

<210> 31
<211> 30
<212> DNA
<213> Artificial Sequence

<220>

<223> Designed oligonucleotide primer to obtain raffinose synthase gene.

<400> 31
agacccgggg aaagctttgg ggttactact 30

<210> 32
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Designed oligonucleotide primer to obtain raffinose synthase gene.

<400> 32
tggatgggaa actttataca ccctgact 28

<210> 33
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Designed oligonucleotide primer to obtain raffinose synthase gene.

<400> 33
gacatgttcc catctacaca cccttgtg 28

<210> 34
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Designed oligonucleotide primer to obtain raffinose synthase gene.

<400> 34
ccaatttatg ttagtgatgt tgttggcaag 30

<210> 35
<211> 26

<212> DNA
<213> Artificial Sequence

<220>
<223> Designed oligonucleotide primer to obtain raffinose synthase gene.

<400> 35
tcgactccca gggtagaatt gtcatc 26

<210> 36
<211> 35
<212> DNA
<213> Artificial Sequence

<220>
<223> Designed oligonucleotide primer to obtain raffinose synthase gene, n is i.

<400> 36
cgattnaang tntggtggac nacncantgg gtngg 35

<210> 37
<211> 38
<212> DNA
<213> Artificial Sequence

<220>
<223> Designed oligonucleotide primer to obtain raffinose synthase gene, n is i, r is a or g.

<400> 37
cantgnacca tntgncancc ntgnarccan tangtncc 38

<210> 38
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Designed oligonucleotide primer to obtain raffinose synthase gene.

<400> 38
gttagggttc atatgaacac cttcaagctc 30

<210> 39
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Designed oligonucleotide primer to obtain raffinose synthase gene.

<400> 39
caacggcgag atcttgcatc gtcaac 26

<210> 40
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Designed oligonucleotide primer to obtain raffinose synthase gene.

<400> 40
ggattcgaca caaaccgcca cgtcatcgtc 30

<210> 41
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Designed oligonucleotide primer to obtain raffinose synthase gene.

<400> 41
ccacgtgcac cacccgaact tatcgac 27

<210> 42
<211> 30
<212> DNA

<213> Artificial Sequence

<220>
<223> Designed oligonucleotide primer to obtain raffinose synthase gene.

<400> 42
aacatcgata ccatcggagt catgtccaat 30

<210> 43
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Designed oligonucleotide primer to obtain raffinose synthase gene.

<400> 43
gttagggttc atatgaacac cttcaagctc 30

<210> 44
<211> 29
<212> DNA
<213> Artificial Sequence

<220>
<223> Designed oligonucleotide primer to obtain raffinose synthase gene.

<400> 44
tctacgtctg gcacgcgctt tgcggctac 29

<210> 45
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Designed oligonucleotide primer to obtain raffinose synthase gene.

<400> 45
gttgacgtca tccacatatt ggagatgttg t 31

<210> 46
<211> 29
<212> DNA
<213> Artificial Sequence

<220>
<223> Designed oligonucleotide primer to obtain raffinose synthase gene.

<400> 46
gttatcgcta gcatggagca ctgtaatga 29

<210> 47
<211> 35
<212> DNA
<213> Artificial Sequence

<220>
<223> Designed oligonucleotide primer to obtain raffinose synthase gene.

<400> 47
aacgagctca atccaaaatc tcatcaaata atcgc 35

<210> 48
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Designed oligonucleotide primer to obtain raffinose synthase gene.

<400> 248
acaatagttg agggcggaag agtag 25

<210> 49
<211> 25
<212> DNA
<213> Artificial Sequence

<220>

<223> Designed oligonucleotide linker to obtain raffinose synthase gene.

<400> 49
gatcgagctc gtgtcggatc cagct 25

<210> 50
<211> 17
<212> DNA
<213> Artificial Sequence

<220>
<223> Designed oligonucleotide linker to obtain raffinose synthase gene.

<400> 50
ggatccgaca cgagctc 17

<210> 51
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Designed oligonucleotide primer to confirm direction of the inserted raffinose synthase gene.

<400> 51
cctcctcgga ttccattgcc cagctatctg 30

<210> 52
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Designed oligonucleotide primer to confirm direction of the inserted raffinose synthase gene.

<400> 52
ggattcgaca caaaccgcca cgtcatcgtc 30

<210> 53
<211> 29
<212> DNA
<213> Artificial Sequence

<220>
<223> Designed oligonucleotide primer to confirm direction of the inserted raffinose synthase gene.

<400> 53
tctacgtctg gcacgcgctt tgcggctac 29

What is claimed is:

1. A raffinose synthase gene which comprises a nucleotide sequence hybridizable with a nucleotide sequence selected from the group consisting of:

(a) a nucleotide sequence encoding the amino acid sequence represented by SEQ ID NO: 1,

(b) the nucleotide sequence represented by SEQ ID NO: 2,

(c) a nucleotide sequence encoding the amino acid sequence represented by SEQ ID NO: 3,

(d) the nucleotide sequence represented by the 236th to 2584th nucleotides in the nucleotide sequence represented by SEQ ID NO: 4,

(e) a nucleotide sequence encoding the amino acid sequence represented by SEQ ID NO: 5,

(f) the nucleotide sequence represented by the 134th to 2467th nucleotides in the nucleotide sequence represented by SEQ ID NO: 6,

(g) a nucleotide sequence encoding the amino acid sequence represented by SEQ ID NO: 7, and

(h) the nucleotide sequence represented by the 1st to 1719th nucleotides in the nucleotide sequence represented by SEQ ID NO: 8,

under stringent conditions, and encoding a protein being capable of binding D-galactosyl group through α (1→6) bond to the hydroxyl group attached to the carbon atom at 6-position of the D-glucose

residue in a sucrose molecule to form raffinose.

2. A raffinose synthase gene comprising a nucleotide sequence encoding the amino acid sequence represented by SEQ ID NO: 1.

3. A raffinose synthase gene comprising the nucleotide sequence represented by SEQ ID NO: 2.

4. A raffinose synthase gene comprising a nucleotide sequence encoding the amino acid sequence represented by SEQ ID NO: 3.

5. A raffinose synthase gene comprising the nucleotide sequence represented by the 236th to 2584th nucleotides in the nucleotide sequence represented by SEQ ID NO: 4.

6. A raffinose synthase gene comprising a nucleotide sequence encoding the amino acid sequence represented by SEQ ID NO: 5.

7. A raffinose synthase gene comprising the nucleotide sequence represented by the 134th to 2467th nucleotides in the nucleotide sequence represented by SEQ ID NO: 6.

8. A raffinose synthase gene comprising a nucleotide sequence encoding the amino acid sequence represented by SEQ ID NO: 7.

9. A raffinose synthase gene comprising the nucleotide sequence represented by the 1st to 1719th nucleotides in the nucleotide sequence represented by SEQ ID NO: 8.

10. A raffinose synthase gene comprising the nucleotide sequence represented by SEQ ID NO: 4, SEQ ID NO: 6 or SEQ ID NO: 8.

11. A nucleic acid comprising a partial nucleotide sequence of the raffinose synthase gene of any one of claims 1 to 10.

12. A method for detecting a nucleic acid containing a raffinose synthase gene which comprises detecting said nucleic acid by hybridization using the labeled nucleic acid of claim 11 as a probe.

13. A method for amplifying a nucleic acid containing a raffinose synthase gene which comprises amplifying said nucleic acid by polymerase chain reaction (PCR) using the nucleic acid of claim 11 as a primer.

14. A method for obtaining a raffinose synthase gene which comprises the steps of:

detecting a nucleic acid containing said raffinose synthase gene by hybridization using the labeled nucleic acid of claim 11 as a probe, and

recovering the detected nucleic acid.

15. A method for obtaining a raffinose synthase gene which comprises the steps of:

amplifying a nucleic acid containing said raffinose synthase gene by PCR using the nucleic acid of claim 11 as a primer, and

recovering the amplified nucleic acid.

16. A nucleic acid comprising a nucleic acid containing the raffinose synthase gene of any one of claims 1 to 10 which is joined to a nucleic acid exhibiting promoter activity in a host cell.

17. A vector comprising the raffinose synthase gene of any one of claims 1 to 10.

18. A transformant, wherein the raffinose synthase gene of any one of claims 1 to 10 is introduced into a host cell.

19. A transformant, wherein the nucleic acid of claim 16 is introduced into a host cell.

20. A transformant, wherein the vector of claim 17 is introduced into a host cell.

21. The transformant of any one of claims 18 to 20, wherein the host is a microorganism.

22. The transformant of any one of claims 18 to 20, wherein the host is a plant.

23. A method for producing a raffinose synthase which comprises the steps of:

culturing or growing the transformant of any one of claims 18 to 22 to produce the raffinose synthase, and

collecting the raffinose synthase.

24. A raffinose synthase comprising the amino acid sequence represented by SEQ ID NO: 1.

25. A raffinose synthase comprising the amino acid sequence represented by SEQ ID NO: 3.

26. A raffinose synthase comprising the amino acid sequence represented by SEQ ID NO: 5.

27. A raffinose synthase comprising the amino acid sequence represented by SEQ ID NO: 7.

Abstract of the disclosure:

A raffinose synthase gene comprising a nucleotide sequence hybridizable with a nucleotide sequence selected from the group consisting of: (a) a nucleotide sequence encoding the amino acid sequence represented by SEQ ID NO: 1, (b) the nucleotide sequence represented by SEQ ID NO: 2, (c) a nucleotide sequence encoding the amino acid sequence of represented by SEQ ID NO: 3, (d) the nucleotide sequence represented by SEQ ID NO: 4 or by the 236th to 2584th nucleotides in the nucleotide sequence represented by SEQ ID NO: 4, (e) a nucleotide sequence encoding the amino acid sequence represented by SEQ ID NO: 5, (f) the nucleotide sequence represented by SEQ ID NO: 6 or by the 134th to 2467th nucleotides in the nucleotide sequence represented by SEQ ID NO: 6, (g) a nucleotide sequence encoding the amino acid sequence represented by SEQ ID NO: 7, and (h) the nucleotide sequence represented by SEQ ID NO: 8 or by the 1st to 1719th nucleotides in the nucleotide sequence represented by SEQ ID NO: 8, under stringent conditions, and encoding a protein being capable of binding D-galactosyl group through α (1→6) bond to the hydroxyl group attached to the carbon atom at 6-position of the D-glucose residue in a sucrose molecule to form raffinose is disclosed.